# Comparison Theorems for Stochastic Chemical Reaction Networks

Felipe A. Campos[1,*], Simone Bruno[2,*], Yi Fu[1], Domitilla Del Vecchio[2], and Ruth J. Williams[1]

[1]*Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0112. Email:* (`fcamposv,yif064,rjwilliams`)`@ucsd.edu`
[2]*Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. Emails:* (`sbruno,ddv`)`@mit.edu`
[*]*These authors contributed equally: F. A. Campos and S. Bruno*

**Abstract.** Continuous-time Markov chains are frequently used as stochastic models for chemical reaction networks, especially in the growing field of systems biology. A fundamental problem for these Stochastic Chemical Reaction Networks (SCRNs) is to understand the dependence of the stochastic behavior of these systems on the chemical reaction rate parameters. Towards solving this problem, in this paper we develop theoretical tools called comparison theorems that provide stochastic ordering results for SCRNs. These theorems give sufficient conditions for monotonic dependence on parameters in these network models, which allow us to obtain, under suitable conditions, information about transient and steady state behavior. These theorems exploit structural properties of SCRNs, beyond those of general continuous-time Markov chains. Furthermore, we derive two theorems to compare stationary distributions and mean first passage times for SCRNs with different parameter values, or with the same parameters and different initial conditions. These tools are developed for SCRNs taking values in a generic (finite or countably infinite) state space and can also be applied for non-mass-action kinetics models. When propensity functions are bounded, our method of proof gives an explicit method for coupling two comparable SCRNs, which can be used to simultaneously simulate their sample paths in a comparable manner. We illustrate our results with applications to models of enzymatic kinetics and epigenetic regulation by chromatin modifications.

**Keywords.** Stochastic chemical reaction networks, monotonicity.

## 1  Introduction

### 1.1  Overview

Stochastic Chemical Reaction Networks (SCRNs) are a class of continuous-time Markov chain models used to describe the stochastic dynamics of a chemical system undergoing a series of reactions which change the numbers of molecules of a finite set of species over time. These models provide a framework for the theoretical study of biochemical systems in areas such as intracellular viral kinetics (see Srivastava et al. [25] and Haseltine & Rawlings [17]),

enzymatic kinetics (see Kang et al. [20] for example) and epigenetic regulation by chromatin modifications (see Bruno et al. [8] for a recently developed model of chromatin regulation).

One of the most interesting questions for biochemical system models is: "What effect does changing reaction rate parameters have on system dynamics?" Indeed, different rate parameters for chemical processes can lead to different stochastic behaviors. One possible approach to evaluate the effect of parameter variations on system dynamics is through comparison theorems for stochastic processes. More precisely, this type of theorem provides inequalities between stochastic processes (see Muller & Stoyan [24] for a general reference on this topic).

In this paper, we employ uniformization and coupling methods (see Grassmann [14] and Keilson [21]) to derive comparison theorems for SCRNs under verifiable sufficient conditions. These theoretical results enable us to develop two novel theorems yielding a direct comparison of mean first passage times and stationary distributions between SCRNs with different rate parameters or initial conditions. We apply these theorems to several examples to illustrate how they can be used to understand how key biological parameters affect stochastic behavior. While a major motivator for our work has been the study of SCRNs, we state our theorems in the context of continuous-time Markov chains, for which the state space is a subset of $\mathbb{Z}_+^d$ (the set of $d$-dimensional vectors with non-negative integer entries), and the set of all possible transition vectors is a finite set. This thereby allows for other applications that have similar characteristics to SCRNs. In addition, for the case of bounded transition intensities satisfying our conditions, we give an explicit concrete coupling of two comparable Markov chains, which can be used to simultaneously simulate them in such a way that their sample paths are monotonically related.

The paper is structured as follows: in Section 2 we introduce some background on stochastic chemical reaction networks needed for this article. We present the main results in Section 3, with proofs provided in Section 5. In Section 4 we apply our theoretical tools to several examples, such as epigenetic regulation by chromatin modifications and enzymatic kinetics. Concluding remarks are presented in Section 6. The Supplementary information (SI) file contains some further details and extensions of the main results and examples in the paper.

## 1.2   Related work

Due to the growing field of systems biology, the mathematical study of chemical reaction networks has seen a wealth of activity lately. Concerning comparison results, considerable work has been conducted on monotonicity properties for deterministic models of chemical reaction networks, i.e., systems of ordinary differential equations describing the dynamics of species concentrations. For example, Angeli et al. [5] proposed a graphical method, based on the monotonicity properties of the reaction rates with respect to species concentrations, to determine global stability properties for the models. More recently, Gori et al. [13] introduced sufficient conditions to verify the existence of a monotonicity property for the concentrations of species for any positive time with respect to their initial concentrations. However, these works do not address how changing parameters affects the behavior of stochastic models.

To the best of our knowledge, no systematic study of stochastic ordering has been conducted for stochastic chemical reaction networks. On a more general level, theorems have been established for stochastic processes and have been specialized for particular classes such as for queueing systems and point processes (see Muller & Stoyan [24] for an introduc-

tion to the topic). For Markov chains, the work of Massey [22] is of special interest, since he establishes criteria for comparison of continuous-time Markov chains in terms of their infinitesimal generators. For relevant work prior to Massey, there is a nice summary in [22]. In particular, Kamae et al. [19] showed that for Markov processes, a comparison between transition probability functions, at all fixed times and for all partially ordered starting points, can be realized in a pathwise stochastic comparison between versions of the Markov processes. In relation to Massey's work, our results provide simplified conditions and extended results for stochastic comparisons, which exploit the structure of stochastic chemical reaction networks. Furthermore, unlike Massey, we do not require a uniform bound on the rates of leaving each state. In addition, under the latter assumption, we explicitly construct versions of the stochastic processes on the same probability space that have comparable sample paths. More detail on the relationship of our work to that of Massey is given in Remark 3.2. In contrast to work on sensitivity analysis of distributions at a finite set of times and which considers only local changes in parameters (see for example Gunawan et al. [15], Gupta & Khammash [16] and references therein), our work provides a sample path comparison between stochastic processes for global changes in their parameters.

## 1.3 Notation and Terminology

Denote by $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ the set of non-negative integers. For an integer $d \geq 1$ we denote by $\mathbb{Z}_+^d$ the set of $d$-dimensional vectors with entries in $\mathbb{Z}_+$. For any integer $d \geq 1$, let $\mathbb{R}^d$ denote the $d$-dimensional Euclidean space. We usually write $\mathbb{R}$ for $\mathbb{R}^1$. We denote by $\mathbb{R}_+^d$ the set of vectors $x \in \mathbb{R}^d$ such that $x_i \geq 0$ for every $1 \leq i \leq d$. For $x \in \mathbb{R}^d$, let $\|x\|_\infty = \sup_{1 \leq i \leq d} |x_i|$ be the supremum norm. In this paper, the sum over the empty set is considered to be 0.

A binary relation $\preccurlyeq$ on a set $\mathcal{X}$ will be called reflexive if $x \preccurlyeq x$ for every $x \in \mathcal{X}$, transitive if $x \preccurlyeq y$ and $y \preccurlyeq z$ implies $x \preccurlyeq z$ for every $x, y, z \in \mathcal{X}$ and antisymmetric if $x \preccurlyeq y$ and $y \preccurlyeq x$ implies $x = y$ for every $x, y \in \mathcal{X}$. A preorder is a binary relation that is reflexive and transitive. A partial order is a preorder that is antisymmetric.

In this paper, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ will consist of a sample space $\Omega$, a $\sigma$-algebra of events $\mathcal{F}$ and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. We will say that two real-valued random variables $Y, Y'$ (defined on possibly different probability spaces) are equal in distribution, denoted as $Y' \overset{\text{dist}}{=} Y$, if their cumulative distribution functions agree. All stochastic processes considered in this paper will have right-continuous sample paths that also have finite left-limits.

# 2 Stochastic Chemical Reaction Networks (SCRNs)

In this section we provide necessary background on Stochastic Chemical Reaction Networks. The reader is referred to Anderson & Kurtz [3] for an introduction to this subject.

We assume there is a finite non-empty set $\mathscr{S} = \{S_1, \ldots, S_d\}$ of $d$ **species**, and a finite non-empty set $\mathscr{R} \subseteq \mathbb{Z}_+^d \times \mathbb{Z}_+^d$ that represents chemical **reactions**. We assume that $(w, w) \notin \mathscr{R}$ for every $w \in \mathbb{Z}_+^d$. The set $\mathscr{S}$ represents $d$ different molecular species in a system subject to reactions $\mathscr{R}$ which change the number of molecules of each species. For each $(v^-, v^+) \in \mathscr{R}$, the $d$-dimensional vector $v^-$ (the **reactant vector**) counts how many molecules of each species are consumed in the reaction, while $v^+$ (the **product vector**) counts how many

molecules of each species are produced. The reaction is usually written as

$$\sum_{i=1}^{d}(v^-)_i \mathrm{S}_i \longrightarrow \sum_{i=1}^{d}(v^+)_i \mathrm{S}_i. \tag{2.1}$$
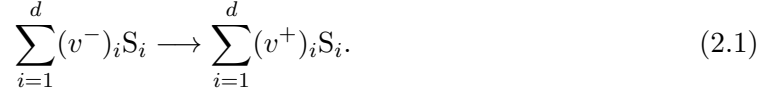
To avoid the use of unnecessary symbols, we will assume that for each $1 \le i \le d$, there exists a vector $w = (a_1, \ldots, a_d)^T \in \mathbb{Z}_+^d$ with $a_i > 0$ such that $(w, v)$ or $(v, w)$ is in $\mathscr{R}$ for some $v \in \mathbb{Z}_+^d$, i.e., each species is either a reactant or a product in some reaction.

The net change in the quantity of molecules of each species due to a reaction $(v^-, v^+) \in \mathscr{R}$ is described by $v^+ - v^-$ and it is called the associated **reaction vector**. We denote the set of reaction vectors $\mathcal{V} := \{v \in \mathbb{Z}^d \,|\, v = v^+ - v^- \text{ for some } (v^-, v^+) \in \mathscr{R}\}$, let $n := |\mathcal{V}|$ the size of $\mathcal{V}$ and enumerate the members of $\mathcal{V}$ as $\{v_1, \ldots, v_n\}$. Note that $\mathcal{V}$ does not contain the zero vector because $\mathscr{R}$ has no elements of the form $(w, w)$. Different reactions might have the same reaction vector. For each $v_j \in \mathcal{V}$ we consider the set $\mathscr{R}_{v_j} := \{(v^-, v^+) \in \mathscr{R} \,|\, v_j = v^+ - v^-\}$. The reaction vectors generate the **stoichiometric subspace** $\mathcal{L} := \mathrm{span}(\mathcal{V})$. For $z \in \mathbb{R}^d$, we call $z + \mathcal{L}$ a **stoichiometric compatibility class**.

Given $(\mathscr{S}, \mathscr{R})$ we will consider an associated continuous-time Markov chain $X = (X_1, \ldots, X_d)$, with a state space $\mathcal{X}$ contained in $\mathbb{Z}_+^d$, which tracks the number of molecules of each species over time. Roughly speaking, the dynamics of $X$ will be given by the following: given a current state $x = (x_1, \ldots, x_d) \in \mathcal{X} \subseteq \mathbb{Z}_+^d$, for each reaction $(v^-, v^+) \in \mathscr{R}$, there is a clock which will ring at an exponentially distributed time (with rate $\Lambda_{(v^-, v^+)}(x)$). The clocks for distinct reactions are independent of one another. If the clock corresponding to $(v^-, v^+) \in \mathscr{R}$ rings first, the system moves from $x$ to $x + v^+ - v^-$ at that time, and then the process repeats. We now define the Markov chain in more detail.

Consider a set of species $\mathscr{S}$ and of reactions $\mathscr{R}$, a set $\mathcal{X} \subseteq \mathbb{Z}_+^d$ and a collection of functions $\{\Lambda_{(v^-, v^+)} : \mathcal{X} \longrightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathscr{R}}$ such that for each $x \in \mathcal{X}$ and $(v^-, v^+) \in \mathscr{R}$, if $x + v^+ - v^- \notin \mathcal{X}$, then $\Lambda_{(v^-, v^+)}(x) = 0$. Now, for $1 \le j \le n$, $v_j \in \mathcal{V}$, define

$$\Upsilon_j(x) := \sum_{(v^-, v^+) \in \mathscr{R}_{v_j}} \Lambda_{(v^-, v^+)}(x). \tag{2.2}$$

Note that for each $x \in \mathcal{X}$ and $1 \le j \le n$, if $x + v_j \notin \mathcal{X}$, then $\Upsilon_j(x) = 0$. A **stochastic chemical reaction network (SCRN)** is a Markov chain $X$ with state space $\mathcal{X}$ and infinitesimal generator[1] $Q$ given for $x, y \in \mathcal{X}$ by

$$Q_{x,y} = \begin{cases} \Upsilon_j(x) & \text{if } y - x = v_j \text{ for some } 1 \le j \le n, \\ -\sum_{j=1}^{n} \Upsilon_j(x) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

The functions $\{\Lambda_{(v^-, v^+)} : \mathcal{X} \longrightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathscr{R}}$ are called **propensity** or **intensity** functions. A common form for the propensity functions is the following associated with **mass action kinetics**:

$$\Lambda_{(v^-, v^+)}(x) = \kappa_{(v^-, v^+)} \prod_{i=1}^{d} (x_i)_{(v^-)_i}, \tag{2.4}$$

---

[1] Note that $Q$ is sometimes called an infinitesimal transition matrix although it may have countably many "rows" and "columns". The entries $Q_{x,y}$ for $x \ne y$ are the infinitesimal transition rates of going from $x$ to $y$: $\mathbb{P}[X(t + h) = y | X(t) = x] = Q_{x,y} h + o(h)$ as $h \to 0$.

4

where $\{\kappa_{(v^-,v^+)}\}_{(v^-,v^+)\in\mathscr{R}}$ are positive constants and for $m, \ell \in \mathbb{Z}_+$, the quantity $(m)_\ell$ is the falling factorial, i.e., $(m)_0 := 1$ and $(m)_\ell := m(m-1)\ldots(m-\ell+1)$.

**Remark 2.1.** Our definition of SCRN allows for some model flexibility. Notice that the propensity functions are not necessarily defined on the whole lattice $\mathbb{Z}_+^d$ and they are not necessarily of the form (2.4). Indeed, in some of our applications, mass-conservation laws restrict the possible values that $X$ may take (see Example 4.4). In addition, there may be other types of kinetics, such as those described by Hill functions (see Example 4.5).

A convenient way to represent such a Markov chain is given in Theorem 6.4.1 of Ethier & Kurtz [12]. For this, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with independent unit rate Poisson processes $N_1, \ldots, N_n$. There is a version of $X$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$X(t) = X(0) + \sum_{j=1}^n v_j N_j \left( \int_0^t \Upsilon_j (X(s)) \, ds \right), \tag{2.5}$$

for every $0 \le t < \tau$, where $\tau$ is the explosion time for $X$ (which may be $+\infty$). From (2.5), it is easy to see that for a SCRN $X$ with initial state $z \in \mathcal{X}$, $X(t)$ will stay in the stoichiometric compatibility class $z + \mathcal{L}$ intersected with $\mathbb{Z}_+^d$ for all time $0 \le t < \tau$, with probability one. For this reason, sometimes it will be convenient to choose $\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^d$, for a fixed $z \in \mathbb{Z}_+^d$.

While our work was initially motivated by questions for SCRNs, we will first develop our results in a more general context of continuous-time Markov chains, for which the state space is contained in $\mathbb{Z}_+^d$ and the set of all possible transition vectors is a finite set, and then illustrate them for SCRNs.

## 3 Main Results

The general stochastic ordering results provided in this paper are relative to a preorder relation on a state space $\mathcal{X} \subseteq \mathbb{Z}_+^d \subseteq \mathbb{R}^d$. We will define the preorder on all of $\mathbb{R}^d$ and then restrict it to various subsets. We introduce this notation and related notation in Section 3.1. In Section 3.2 we present the main results of this article, and in Section 3.3 we discuss relevant consequences for the comparison of (mean) first passage times and stationary distributions.

### 3.1 Preorders in $\mathbb{R}^d$

Let $m, d \ge 1$ be integers. Denote by $\le$ the usual componentwise partial order on $\mathbb{R}^d$, i.e., for $x, y \in \mathbb{R}^d$, $x \le y$ whenever $x_i \le y_i$ for every $1 \le i \le d$. Additionally, we write $x < y$ whenever $x_i < y_i$ for every $1 \le i \le d$. For the rest of the paper, we consider a matrix $A \in \mathbb{R}^{m \times d}$, where no row of $A$ is identically zero.

**Definition 3.1.** For $x, y \in \mathbb{R}^d$, we say that $x \preccurlyeq_A y$ whenever $A(y - x) \ge 0$.

For the matrix $A$, consider the convex cone $K_A := \{x \in \mathbb{R}^d \mid Ax \ge 0\}$. Note that $x \preccurlyeq_A y$ holds if and only if $y - x \in K_A$. Moreover, the relation $\preccurlyeq_A$ is reflexive and transitive, and therefore a preorder on $\mathbb{R}^d$. Also, for this relation,

$$\text{if } x \preccurlyeq_A y, \text{ then } x + z \preccurlyeq_A y + z \text{ for any } z \in \mathbb{R}^d. \tag{3.1}$$

5

For any $x \in \mathbb{R}^d$ consider the set

$$K_A + x = \{y \in \mathbb{R}^d \mid A(y - x) \geq 0\} = \{y \in \mathbb{R}^d \mid x \preccurlyeq_A y\}.$$

In the coming sections, we will consider the notions of increasing and decreasing sets with respect to $\preccurlyeq_A$ in a given subset of $\mathbb{Z}_+^d$. More concretely, consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$. We will say that a set $\Gamma \subseteq \mathcal{X}$ is **increasing** in $\mathcal{X}$ with respect to $\preccurlyeq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $x \preccurlyeq_A y$ implies that $y \in \Gamma$. We observe that, for $x \in \mathcal{X}$, the set

$$(K_A + x) \cap \mathcal{X} = \{y \in \mathcal{X} \mid x \preccurlyeq_A y\} \tag{3.2}$$

is increasing in $\mathcal{X}$ by the transitivity property of $\preccurlyeq_A$. On the other hand, we will say that a set $\Gamma \subseteq \mathcal{X}$ is **decreasing** in $\mathcal{X}$ with respect to $\preccurlyeq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $y \preccurlyeq_A x$ implies that $y \in \Gamma$. We will say that a point $x$ is **maximal** (resp. **minimal**) in $\mathcal{X}$ if for every $y \in \mathcal{X}$, $x \preccurlyeq_A y$ (resp. $y \preccurlyeq_A x$) implies that $x = y$. In this case, the set $\Gamma = \{x\}$ would be increasing (resp. decreasing) in $\mathcal{X}$.

**Remark 3.1.** If $\operatorname{rank}(A) = d$, then the relation $\preccurlyeq_A$ will be antisymmetric and therefore a partial order on $\mathbb{R}^d$. Indeed, if $\operatorname{rank}(A) = d$, then $A(y - x) = 0$ implies that $x = y$. In addition, $\preccurlyeq_A$ will then be a partial order when restricted to $\mathcal{X} \subset \mathbb{Z}_+^d$. Throughout this article, we will not assume that $\operatorname{rank}(A) = d$ and therefore, the relation $\preccurlyeq_A$ might not be a partial order on $\mathcal{X}$ (see Examples 4.1, 4.2, and 4.3).

## 3.2 Stochastic Comparison Theorems

The fundamental objects in the following results are a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$ and a pair of continuous-time Markov chains $X$ and $\check{X}$ with the same state space $\mathcal{X}$ and where it is assumed that the set of all possible transition vectors for $X$ or $\check{X}$ is a finite set. We denote the size of this set by $n$. A primary example of this setup is two stochastic chemical reaction networks as described in Section 2 with different propensity functions. We will now formally introduce the notation for stating our results.

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, an integer $n \geq 1$ and a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$, where $0$ is the origin in $\mathbb{Z}^d$. Consider two collections of functions $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n)$ defined on $\mathcal{X}$ and taking values in $\mathbb{R}_+$, such that for every $1 \leq j \leq n$ and $x \in \mathcal{X}$:

$$\text{if } x + v_j \notin \mathcal{X}, \text{ then } \Upsilon_j(x) = \check{\Upsilon}_j(x) = 0. \tag{3.3}$$

Consider a continuous-time Markov chain $X$ on the state space $\mathcal{X}$ with infinitesimal generator $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$ defined for $x, y \in \mathcal{X}$ by

$$Q_{x,y} := \begin{cases} \Upsilon_j(x) & \text{if } y - x = v_j \text{ for some } 1 \leq j \leq n, \\ -\sum_{j=1}^n \Upsilon_j(x) & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

Consider the analogous continuous-time Markov chain $\check{X}$ with infinitesimal generator $\check{Q}$ as in (3.4) but with functions $\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n$ instead of $\Upsilon_1, \ldots, \Upsilon_n$. We call $X$ and $\check{X}$ the continuous-time Markov chains associated with $\Upsilon$ and $\check{\Upsilon}$ respectively. We will assume that $X$ and $\check{X}$ do not explode in finite time. The following is our main result.
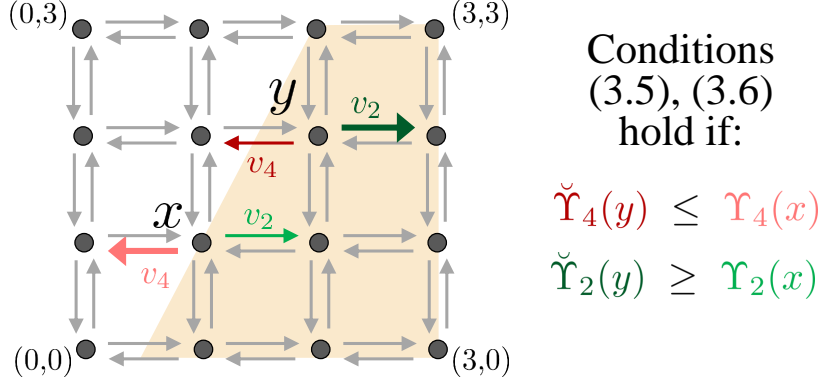
6

Figure 1: **Pictorial representation of conditions (3.5), (3.6) for a certain $(K_A + x) \cap \mathcal{X}$ in a two-dimensional lattice.** Here, $\mathcal{X} = \{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$, $n = 4$, $v_1 = (0, 1)^T$, $v_2 = (1, 0)^T$, $v_3 = (0, -1)^T$, $v_4 = (-1, 0)^T$, where $T$ denotes transpose, $A = [2 \ -1]$, and $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid [2 \ -1](w - x) \geq 0\}$. In the graph, $(K_A + x) \cap \mathcal{X}$ consists of the states (black dots) that lie in the light orange region and the arrows represent possible transitions along $v_1, v_2, v_3, v_4$ between states. For the exhibited states $x, y \in \mathcal{X}$ with $x \preccurlyeq_A y$, the light green (dark green) and light red (dark red) arrows represent the transitions with rates $\Upsilon_2(x)$ ($\check{\Upsilon}_2(y)$) and $\Upsilon_4(x)$ ($\check{\Upsilon}_4(y)$) for the Markov chain $X$ ($\check{X}$). Higher transitions rates are associated with thicker arrows. To check the conditions (3.5) and (3.6), since $y + v_4 \notin K_A + x$ and $y \notin K_A + x + v_2$, we need to check that $\check{\Upsilon}_4(y) \leq \Upsilon_4(x)$ and $\check{\Upsilon}_2(y) \geq \Upsilon_2(x)$.

**Theorem 3.1.** *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n)$, such that (3.3) holds and the associated continuous-time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{R}^{m \times d}$ with non-zero rows and suppose that for every $x, y \in \mathcal{X}$ such that $x \preccurlyeq_A y$ the following hold:*

$$\check{\Upsilon}_j(y) \leq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } y + v_j \in \mathcal{X} \setminus (K_A + x), \qquad (3.5)$$

*and*

$$\check{\Upsilon}_j(y) \geq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } x + v_j \in \mathcal{X} \text{ and } y \notin K_A + x + v_j. \qquad (3.6)$$

*Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous-time Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators $Q$ and $\check{Q}$, associated with $\Upsilon$ and $\check{\Upsilon}$, respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$ and such that:*

$$\mathbb{P}\left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0\right] = 1. \qquad (3.7)$$

An example of checking conditions (3.5) and (3.6) is given in Figure 1. The proof of Theorem 3.1 is given in Section 5.1. The main idea in the construction of the processes $X$ and $\check{X}$ is *uniformization* (see Chapter 2 in Keilson [21]) together with a suitable coupling. Our proof uses a single Poisson process together with a sequence of i.i.d. uniform random variables to determine potential jumps for the two continuous-time Markov chains, where for $X$ and $\check{X}$, potential jumps in the same direction $v_j$ are coupled together, and their probabilities of acceptance are given by normalized versions of their infinitesimal transition rates

$\Upsilon_j$ and $\breve{\Upsilon}_j$. Uniformization can be done provided the diagonal terms of the infinitesimal generators are uniformly bounded in size. In the proof of Theorem 3.1, we initially make this assumption on $Q$ and $\breve{Q}$ in order to construct $X$ and $\breve{X}$. We then generalize the result to Markov chains that do not explode in finite time by using a truncation and limiting procedure. The construction mentioned here, for the case where the diagonal terms of the infinitesimal generators are uniformly bounded in size, besides playing a key role in our proofs, is also the basis for an algorithm described in Section S.4, which provides a way to simultaneously simulate the processes $X$ and $\breve{X}$ in a comparable manner.

**Remark 3.2.** In Theorem 5.3 of Massey [22], the author provides a necessary and sufficient condition for stochastic comparison of continuous-time Markov chains at each fixed time for all partially ordered initial conditions. By the work of Kamae et al. [19], the conditions in [22] imply the existence of a coupling of continuous-time Markov chains so that a relation such as (3.7) holds. Massey's condition requires that $\sum_{w \in \Gamma} Q_{xw} \leq \sum_{w \in \Gamma} \breve{Q}_{yw}$ for every $x \preccurlyeq_A y$ and every set $\Gamma \subseteq \mathcal{X}$ that is increasing in $\mathcal{X}$ with respect to $\preccurlyeq_A$ and such that either $x \in \Gamma$ or $y \notin \Gamma$. These inequalities can often be difficult to check since first, they involve computing sums of terms in the infinitesimal generators and second, the form of all increasing sets can be hard to determine. In Theorem 3.1 we overcome these obstacles by providing simplified sufficient conditions that involve only pointwise comparison of entries in the infinitesimal generators associated to each of the transition vectors $v_j$. Besides this practical value, in our context, our results go beyond the work of Massey [22], since he assumes that $\preccurlyeq_A$ is a partial order (we only assume preorder) and he assumes that the diagonal entries of the infinitesimal generators are bounded (we generalize to non-exploding Markov chains). Our proof has a commonality with the work of Massey in the sense that we also use uniformization. It is different in the sense that, when infinitesimal transition rates are bounded, we construct an explicit coupling for all time, exploiting the simplified nature of our conditions, while Massey does not provide an explicit coupling. Instead, he proves existence of a stochastic comparison for each fixed time, using a semigroup approach.

Conditions (3.5) and (3.6) may be simplified if we consider a particular relation between the matrix $A$ and the vectors $v_1, \ldots, v_n$ in which $A \in \mathbb{Z}^{m \times d}$ and $Av_j$ has entries taking values only in $\{-1, 0, 1\}$ for every $1 \leq j \leq n$. More concretely, let us consider a class of continuous-time Markov chains such that, for a given matrix $A$ with non-zero rows, if the Markov chain starts within the set $K_A + x$, then to go outside of it, the process will necessarily hit its boundary. In this case, we can derive a theorem whose conditions must be checked only on the boundary of $K_A + x$ because the only transitions that can lead the Markov chain outside or inside the set $K_A + x$ are ones starting on the boundary of $K_A + x$. Before stating the theorem, let us introduce the sets $\partial_i(K_A + x) := \{y \in K_A + x \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle\}$ [2] for

---

[2]Here, for convenience of notation, let $A_{i\bullet}$ denote the row vector corresponding to the $i$-th row of $A$, for $1 \leq i \leq m$. In this article we will adopt the convention of considering the inner product $\langle \cdot, \cdot \rangle$ as a function of a row vector in its first entry and as a function of a column vector in the second entry. In particular, $\langle A_{i\bullet}, x \rangle = \sum_{k=1}^{d} A_{ik} x_k$.

each $1 \le i \le m$. We can then characterize[3] the boundary of $K_A + x$ as follows:

$$\partial(K_A + x) = \bigcup_{i=1}^{m} \partial_i(K_A + x). \tag{3.8}$$

**Theorem 3.2.** *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n)$ such that (3.3) holds and the associated continuous-time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows and suppose that both of the following conditions hold:*

(i) *For each $1 \le j \le n$, the vector $Av_j$ has entries in $\{-1, 0, 1\}$ only.*

(ii) *For each $x \in \mathcal{X}$, $1 \le i \le m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that*

$$\check{\Upsilon}_j(y) \le \Upsilon_j(x), \quad \text{for each } 1 \le j \le n \text{ such that } \langle A_{i\bullet}, v_j \rangle < 0, \tag{3.9}$$

*and*

$$\check{\Upsilon}_j(y) \ge \Upsilon_j(x), \quad \text{for each } 1 \le j \le n \text{ such that } \langle A_{i\bullet}, v_j \rangle > 0. \tag{3.10}$$

*Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous-time Markov chains $X = \{X(t), t \ge 0\}$ and $\check{X} = \{\check{X}(t), t \ge 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators given by $Q$ and $\check{Q}$, associated with $\Upsilon$ and $\check{\Upsilon}$ respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$, and such that:*

$$\mathbb{P}\left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \ge 0\right] = 1. \tag{3.11}$$

The proof of this theorem is given in Section 5.2 and involves checking that (3.5) and (3.6) of Theorem 3.1 hold, using conditions *(i)* and *(ii)* of Theorem 3.2.

**Remark 3.3.** In the context of Theorem 3.2, it is possible that for $x \in \mathcal{X}$, and $y \in \partial_{i_1}(K_A + x) \cap \partial_{i_2}(K_A + x) \cap \mathcal{X}$ with $i_1 \ne i_2$, it happens that $\langle A_{i_1\bullet}, v_j \rangle < 0$ and $\langle A_{i_2\bullet}, v_j \rangle > 0$ for some $1 \le j \le n$. For condition *(ii)* to hold, we must then have $\check{\Upsilon}_j(y) = \Upsilon_j(x)$.

When there are multiple vectors $v_j$ with a common value for $Av_j$, the pointwise comparison in $j$, for $1 \le j \le n$, in conditions (3.9) and (3.10) in Theorem 3.2, can be weakened. To this end, let us introduce the set of distinct vectors $\{\eta^1, \ldots, \eta^s\}$ formed by $Av_j$, for $1 \le j \le n$, where $s$ denotes the cardinality of this set. Consider the subsets of indices

$$G^k := \{j \mid 1 \le j \le n \text{ and } Av_j = \eta^k\}, \quad \text{for } 1 \le k \le s. \tag{3.12}$$

Then we have the following theorem.

---

[3]The fact that $A$ does not contain zero rows allows for equation (3.8) to hold. In fact, let $A \in \mathbb{R}^{m \times d}$ be a matrix that is not identically zero and let $\bar{A}$ be the matrix obtained from $A$ by erasing any rows that contain all zeros. Then, for $x, y \in \mathbb{R}^d$, $A(y - x) \ge 0$ if and only if $\bar{A}(y - x) \ge 0$, and so $K_A + x = K_{\bar{A}} + x$ and $x \preccurlyeq_A y$ if and only if $x \preccurlyeq_{\bar{A}} y$. However, if $A$ contains a row $A_{i\bullet}$ such that $A_{i\bullet} = 0$, then $\partial_i(K_A + x) = K_A + x$ and if $K_A + x$ has nonempty interior, then equation (3.8) will not hold. Consequently, we have made the assumption that $A$ has no zero rows.

**Theorem 3.3.** *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n)$ such that (3.3) holds and the associated continuous-time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows and suppose that both of the following conditions hold:*

*(i) For each $1 \leq j \leq n$, the vector $Av_j$ has entries in $\{-1, 0, 1\}$ only.*

*(ii) For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that*

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \leq \sum_{j \in G^k} \Upsilon_j(x), \quad \textit{for each } k \textit{ such that } \eta_i^k < 0, \tag{3.13}$$

*and*

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \geq \sum_{j \in G^k} \Upsilon_j(x), \quad \textit{for each } k \textit{ such that } \eta_i^k > 0. \tag{3.14}$$

*Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preceq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous-time Markov chains $X = \{X(t), \, t \geq 0\}$ and $\check{X} = \{\check{X}(t), \, t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators $Q$ and $\check{Q}$, associated with $\Upsilon$ and $\check{\Upsilon}$ respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$ and such that:*

$$\mathbb{P}\left[X(t) \preceq_A \check{X}(t) \textit{ for every } t \geq 0\right] = 1. \tag{3.15}$$

The proof of this theorem is given in Section 5.3.

**Remark 3.4.** If $\Upsilon = \check{\Upsilon}$, Theorems 3.1, 3.2 and 3.3 give sufficient conditions for monotonic dependence of the stochastic dynamic behavior on the initial condition. In the sense of Massey [22], this notion corresponds to constructing a *strongly monotone Markov chain*.

**Remark 3.5.** For *deterministic* dynamical systems, there is a considerable literature giving monotonicity conditions with respect to initial conditions (see e.g., Hirsch & Smith [18]). Furthermore, Angeli & Sontag [4] extended the concept of monotone systems to systems having external inputs (i.e., $\dot{x} = f(x, u)$, with $x$ representing the state and $u$ representing the input). More precisely, they developed tools to prove monotonic dependence of the deterministic dynamic behavior on the initial condition and external input, provided that certain sign conditions on the first partial derivatives of the function $f(x, u)$ are satisfied on the entire state and input space. These theoretical tools can be used also to study how changing a system parameter affects the deterministic behavior of the system, by viewing $u$ as the system parameter of interest.

**Remark 3.6.** Checking the conditions in Theorems 3.2 and 3.3 (if they hold) is less cumbersome than checking the conditions in Theorem 3.1. In fact, compared to Theorem 3.1, for Theorems 3.2 and 3.3, the conditions must be checked only on the boundaries of $K_A + x$, given that condition $(i)$ there is assumed to hold. Furthermore, Theorem 3.3 has less restrictive conditions (i.e., comparing sums of infinitesimal rates associated with transitions inwards or outwards with respect to the hyperplanes $\{z \in \mathbb{R}^d \,|\, \langle A_{i\bullet}, z \rangle = \langle A_{i\bullet}, x \rangle = \langle A_{i\bullet}, y \rangle\}$, $1 \leq i \leq m$, instead of comparing transition rates one-by-one for $1 \leq j \leq n$).

### 3.3 Monotonicity properties for (Mean) First Passage Times and Stationary Distributions

The first consequence of our main results is for first passage times and it is related to stochastic orderings of real-valued random variables. Let $Y$ and $Z$ be real-valued random variables with cumulative distribution functions $F_Y$ and $F_Z$ respectively. We say that $Y$ is smaller than $Z$ in the **usual stochastic order**, written $Y \preccurlyeq_{st} Z$ if $F_Y(t) \geq F_Z(t)$ for every $t \in \mathbb{R}$. The relation $Y \preccurlyeq_{st} Z$ is equivalent to the existence of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables $Y' \stackrel{\text{dist}}{=} Y$ and $Z' \stackrel{\text{dist}}{=} Z$ defined there such that $\mathbb{P}(Y' \leq Z') = 1$. Furthermore, it is equivalent to the condition: $\int_{-\infty}^{\infty} f(x) dF_Y(x) \leq \int_{-\infty}^{\infty} f(x) dF_Z(x)$ for every bounded increasing function $f : \mathbb{R} \longrightarrow \mathbb{R}$. The reader may consult Chapter 1 in Muller & Stoyan [24] for the corresponding proofs and further properties of this notion.

**Theorem 3.4.** *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n)$, such that (3.3) holds and the associated continuous-time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{R}^{m \times d}$ with non-zero rows and suppose that at least one of the following holds:*

*(i) For every $x, y \in \mathcal{X}$ such that $x \preccurlyeq_A y$, conditions (3.5) and (3.6) are satisfied.*

*(ii) The matrix $A$ has integer-valued entries and conditions (i) and (ii) in Theorem 3.2 are satisfied.*

*(iii) The matrix $A$ has integer-valued entries and conditions (i) and (ii) in Theorem 3.3 are satisfied.*

*Let $x^\circ, \check{x}^\circ \in \mathcal{X}$ be such that $x^\circ \preccurlyeq_A \check{x}^\circ$ and let $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ be two continuous-time Markov chains (possibly defined on different probability spaces), each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators $Q$ and $\check{Q}$, associated with $\Upsilon$ and $\check{\Upsilon}$ respectively, and with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$. For a non-empty set $\Gamma \subseteq \mathcal{X}$, consider $T_\Gamma := \inf\{t \geq 0 \mid X(t) \in \Gamma\}$ and $\check{T}_\Gamma := \inf\{t \geq 0 \mid \check{X}(t) \in \Gamma\}$. If $\Gamma$ is increasing in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$, then*

$$\check{T}_\Gamma \preccurlyeq_{st} T_\Gamma, \tag{3.16}$$

*and the mean first passage time of $\check{X}$ from $\check{x}^\circ$ to $\Gamma$ is dominated by the mean first passage time of $X$ from $x^\circ$ to $\Gamma$. If $\Gamma$ is decreasing in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$, then*

$$T_\Gamma \preccurlyeq_{st} \check{T}_\Gamma, \tag{3.17}$$

*and the mean first passage time of $X$ from $x^\circ$ to $\Gamma$ is dominated by the mean first passage time of $\check{X}$ from $\check{x}^\circ$ to $\Gamma$.*

*Proof.* By Theorem 3.1, 3.2 or 3.3, we can construct two versions of the processes $X$ and $\check{X}$ on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with initial conditions $x^\circ$ and $\check{x}^\circ$, respectively and such that (3.7) or (3.11) or (3.15) hold. We denote these versions again by $X$ and $\check{X}$, and we observe that to show (3.16), it suffices to show that for an increasing set $\Gamma$, $\mathbb{P}[\check{T}_\Gamma \leq T_\Gamma] = 1$ for $T_\Gamma$ and $\check{T}_\Gamma$ associated with these versions of $X$ and $\check{X}$. To see that this holds, let $\tilde{\Omega}$ be a set of probability one on which

$$X(t) \preccurlyeq_A \check{X}(t), \quad \text{for all } t \geq 0 \tag{3.18}$$

(this exists by (3.7), (3.11) or (3.15)). On $\{T_\Gamma = +\infty\}$, it is clear that $\breve{T}_\Gamma \leq T_\Gamma$. For each $\omega \in \{T_\Gamma < +\infty\} \cap \tilde{\Omega}$ and $\varepsilon > 0$ there is $\tau_\varepsilon(\omega) \in [T_\Gamma(\omega), T_\Gamma(\omega) + \varepsilon)$ such that $X(\tau_\varepsilon(\omega)) \in \Gamma$ and by (3.18), $X(\tau_\varepsilon(\omega)) \preccurlyeq_A \breve{X}(\tau_\varepsilon(\omega))$. And then, since $\Gamma$ is increasing, $\breve{X}(\tau_\varepsilon(\omega)) \in \Gamma$. It follows that $\breve{T}_\Gamma(\omega) \leq T_\Gamma(\omega) + \varepsilon$ and letting $\varepsilon \to 0$ we obtain that $\breve{T}_\Gamma(\omega) \leq T_\Gamma(\omega)$. It follows that $\mathbb{P}[\breve{T}_\Gamma \leq T_\Gamma] = 1$. For the result on mean first passage times, let $\overline{F}_{T_\Gamma} := 1 - F_{T_\Gamma}$ and $\overline{F}_{\breve{T}_\Gamma} := 1 - F_{\breve{T}_\Gamma}$ represent the complementary cumulative distribution functions for $T_\Gamma$ and $\breve{T}_\Gamma$, respectively. Observe that (3.16) implies that $\overline{F}_{\breve{T}_\Gamma} \leq \overline{F}_{T_\Gamma}$. For a non-negative random variable, the mean of the random variable is given by the Lebesgue integral of the complementary cumulative distribution function. Consequently, the mean first passage time for $\breve{X}$ from $\breve{x}^\circ$ to $\Gamma$ is given by $\int_0^\infty \overline{F}_{\breve{T}_\Gamma}(t) dt \leq \int_0^\infty \overline{F}_{T_\Gamma}(t) dt$, where the latter is the mean first passage time for $X$ from $x^\circ$ to $\Gamma$. If $\Gamma$ is decreasing, analogous arguments yield the results stated for that case. $\square$

The second consequence of our results provides a comparison result for stationary distributions.

**Theorem 3.5.** *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\breve{\Upsilon} = (\breve{\Upsilon}_1, \ldots, \breve{\Upsilon}_n)$, such that (3.3) holds and the associated continuous-time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{R}^{m \times d}$ with non-zero rows and suppose that at least one of the following holds:*

*(i) For every $x, y \in \mathcal{X}$ such that $x \preccurlyeq_A y$, conditions (3.5) and (3.6) are satisfied.*

*(ii) The matrix $A$ has integer-valued entries and conditions (i) and (ii) in Theorem 3.2 are satisfied.*

*(iii) The matrix $A$ has integer-valued entries and conditions (i) and (ii) in Theorem 3.3 are satisfied.*

*Assume that the two continuous-time Markov chains on the set $\mathcal{X}$ with infinitesimal generators $Q$ and $\breve{Q}$, associated with $\Upsilon$ and $\breve{\Upsilon}$ respectively, are irreducible and positive recurrent on $\mathcal{X}$, and denote the associated stationary distributions by $\pi$ and $\breve{\pi}$, respectively. If $\Gamma \subseteq \mathcal{X}$ is a non-empty set that is increasing in $\mathcal{X}$ with respect to $\preccurlyeq_A$, then*

$$\sum_{x \in \Gamma} \pi_x \leq \sum_{x \in \Gamma} \breve{\pi}_x. \tag{3.19}$$

*If $\Gamma \subseteq \mathcal{X}$ is a non-empty set that is decreasing in $\mathcal{X}$ with respect to $\preccurlyeq_A$, then*

$$\sum_{x \in \Gamma} \breve{\pi}_x \leq \sum_{x \in \Gamma} \pi_x. \tag{3.20}$$

*Proof.* As in the proof of Theorem 3.4, we can construct two versions of the processes $X$ and $\breve{X}$ on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for some pair of initial conditions $x^\circ \preccurlyeq_A \breve{x}^\circ$. If $\Gamma \subseteq \mathcal{X}$ is increasing, equation (3.7) or (3.11) or (3.15) yields that $\mathbb{P}(X(t) \in \Gamma) \leq \mathbb{P}(\breve{X}(t) \in \Gamma)$ for every $t \geq 0$. By letting $t \to \infty$ and observing that the stationary distribution is the steady-state distribution under our assumptions of irreducibility and positive recurrence, we obtain (3.19). If $\Gamma$ is decreasing, an analogous argument yields (3.20). $\square$

**Remark 3.7.** A special case of Theorems 3.4 and 3.5 is when $\Gamma = \{x\}$ for some maximal or minimal element $x \in \mathcal{X}$.

In the next section, we give examples which illustrate Theorem 3.2 (see Examples 4.1, 4.2, 4.4 and 4.5), Theorem 3.3 (see Example 4.3), Theorem 3.4 and Theorem 3.5 for continuous-time Markov chains that are stochastic chemical reaction networks. For Examples 4.1, 4.2 and 4.3, the state space $\mathcal{X}$ will be a stoichiometric compatibility class $z + \mathcal{L}$ intersected with $\mathbb{Z}_+^d$. For Examples 4.4 and 4.5, we work with reduced Markov chains and the state space $\mathcal{X}$ will be a projection of a suitable higher dimensional stoichiometric compatibility class $z + \mathcal{L}$ intersected with $\mathbb{Z}_+^d$.

## 4   Examples

In this section, we apply the theoretical tools developed in the paper to several examples. While in Examples 4.1, 4.3 and 4.4 the Markov chains analyzed have a finite state space, in Examples 4.2 and 4.5 the Markov chains have a countably infinite state space, but it is straightforward to verify that they do not explode (see SI - Sections S.1.2 and S.1.3, respectively). The choice of matrix $A$ in each example is based on the specific monotonicity relationship of interest. While for simpler cases the choice of $A$ is straightforward, for more complicated systems the choice can be more subtle. In many cases, in order to study the monotonicity properties for the stochastic behavior of our system, we can rely on Theorem 3.2, which provides a reasonable approach to narrow down the choices for suitable $A$. The approach consists in solving, for each row $i$, the system of equations $\sum_{k=1}^{d} A_{ik}(v_j)_k = b_{ij}$, with $b_{ij}$ equal to $1, -1$, or $0$ depending, based on the monotonicity relationship of interest, whether we expect that the Markov chain transition in the direction $v_j$ leads inside, outside, or is parallel to the boundary of the region $K_A + x$. Finally, it is worth noticing that, while all the following examples compare two identical reaction networks with different rate constants, our theory can also be applied to compare two different reaction networks as long as they have the same net reaction vectors $\{v_j\}_{j=1}^{n}$.

**Example 4.1. Enzyme kinetics I**
Let us consider a classic model of enzyme kinetics (see Michaelis & Menten [23] and Kang et al. [20]), where an enzyme catalyzes the conversion of a substrate to a product. The species considered here are substrate (S), enzyme (E), intermediate enzyme-substrate complex (SE), and product (P), and the chemical reaction system is depicted in Fig. 2(a). We are interested in how the rate constant $\kappa_3$ affects the time to convert the substrate to the final product.

To this end, let us first introduce the set of species $\mathscr{S} = \{S, P, E, SE\}$, and the set of reactions $\mathscr{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+)\}$, where $v_1^- = v_2^+ = (1, 0, 1, 0)^T$, $v_1^+ = v_2^- = v_3^- = (0, 0, 0, 1)^T$, $v_3^+ = (0, 1, 1, 0)^T$, where $T$ denotes transpose. At a given time, let the counts of each of the species S, P, E and SE be denoted by $n_S, n_P, n_E$ and $n_{SE}$, respectively. The state of the associated Markov chain is $(n_S, n_P, n_E, n_{SE})$. The potential transitions of the Markov chain are in three possible directions:

$$v_1 = v_1^+ - v_1^- = (-1, 0, -1, 1)^T, v_2 = v_2^+ - v_2^- = (1, 0, 1, -1)^T, v_3 = v_3^+ - v_3^- = (0, 1, 1, -1)^T.$$

Fixing integers $S_{tot}, E_{tot} > 0$, we have a stoichiometric compatibility class $z + \mathcal{L}$ with $z = (S_{tot}, 0, E_{tot}, 0)$ and $\mathcal{L} := \text{span}\{v_1, v_2, v_3\}$, which is contained in a two-dimensional
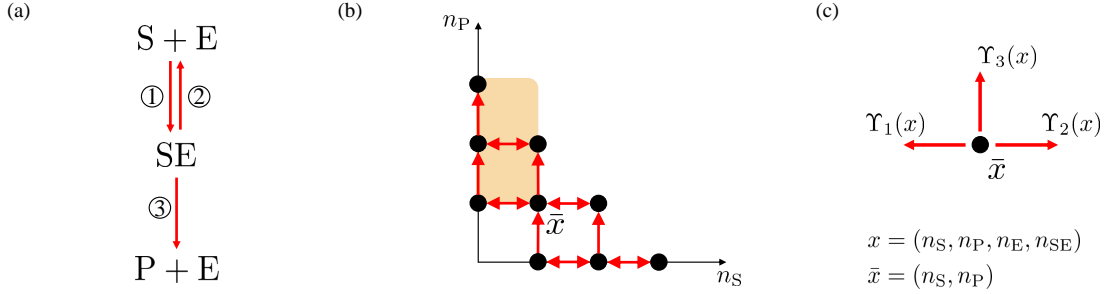
13

Figure 2: **Reaction model and corresponding Markov chain for enzymatic kinetics I example.** (a) Chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Projected Markov chain graph for one stoichiometric compatibility class with two conservation laws $n_S + n_P + n_{SE} = S_{tot} = 3$ and $n_E + n_{SE} = E_{tot} = 2$. The projection takes a state $x = (n_S, n_P, n_E, n_{SE}) = (n_S, n_P, E_{tot} - S_{tot} + n_S + n_P, S_{tot} - n_S - n_P)$ to $\bar{x} = (n_S, n_P)$. We use black dots to represent the states, red double-ended arrows to represent transitions in both directions and red single-ended arrows to represent transitions in one direction. Note that $\bar{x} = (0,0)$ is not a vertex in the graph because $0 \le n_E = E_{tot} - S_{tot} + n_S + n_P$, and so $n_S + n_P \ge 3 - 2 = 1$. We use orange to highlight the projection of the region $K_A + x$ intersected with the stoichiometric compatibility class, where $A$ is defined in (4.2). (c) The projections of the directions of the possible transitions of the Markov chain. The transition rates $\Upsilon_1(x)$, $\Upsilon_2(x)$, and $\Upsilon_3(x)$ are defined in (4.1).

affine subspace of four dimensional space. Then, the state space of the Markov chain is

$$\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}_+^4 | x_1 + x_2 + x_4 = S_{tot}, x_3 + x_4 = E_{tot}\}.$$

The two constraints described in the last expression for $\mathcal{X}$ characterize the two linearly independent conservation laws for this chemical reaction system: $n_S + n_P + n_{SE} = S_{tot}$ and $n_E + n_{SE} = E_{tot}$.

Given a state $x = (x_1, x_2, x_3, x_4) \in \mathcal{X}$, following mass-action kinetics, the infinitesimal transition rates are

$$\Upsilon_1(x) = \kappa_1 x_1 x_3, \quad \Upsilon_2(x) = \kappa_2 x_4, \quad \Upsilon_3(x) = \kappa_3 x_4, \tag{4.1}$$

for constants $\kappa_1, \kappa_2, \kappa_3 > 0$. Here, we have used $\kappa_j$ as an abbreviation for $\kappa_{(v_j^-, v_j^+)}, j = 1, 2, 3$. We will use similar abbreviations in the other examples too.

We note that the projected process $(X_1, X_2)(\cdot)$ is still a continuous-time Markov chain, and we could apply our theory to it. However, when the functions $\Upsilon_j, j = 1, 2, 3$, are written in terms of these two components, they will have a more complex, non-mass action form. Here we apply our theory directly to our four dimensional Markov chain. For the purpose of visualization, Fig. 2(b) shows the two dimensional projection of the four dimensional Markov chain graph for one stoichiometric compatibility class. In Examples 4.2 and 4.3, we also analyze Markov chains without projections, and in Examples 4.4 and 4.5, we analyze projected Markov chains.

In order to study how the rate constant $\kappa_3$ affects the time to convert the substrate to the final product, let us define the state $(0, S_{tot}, E_{tot}, 0)$ associated with $n_P = S_{tot}$ as $p$, the state $(S_{tot}, 0, E_{tot}, 0)$ associated with $n_S = S_{tot}$ as $s$, and the mean first passage time to reach the state $p$, starting from $s$, as $\mathbb{E}_s[T_p]$. We will verify that the assumptions of Theorems 3.2, 3.4

14

hold and exploit them to determine how $\kappa_3$ affects $\mathbb{E}_s[T_p]$. To this end, define the matrix

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{4.2}$$

and consider the preorder $x \preccurlyeq_A y$, defined by $A(y - x) \geq 0$, and the set $K_A + x = \{w \in \mathbb{R}^4 \mid x \preccurlyeq_A w\}$. Let us also consider the infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x)$ and $\check{\Upsilon}_3(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x)$ and $\Upsilon_3(x)$, but with $\check{\kappa}_1 = \kappa_1$, $\check{\kappa}_2 = \kappa_2$, $\check{\kappa}_3 > \kappa_3$ in place of $\kappa_1$, $\kappa_2$, $\kappa_3$, respectively. Condition $(i)$ of Theorem 3.2 (i.e., for every $1 \leq j \leq n$, the vector $Av_j$ has entries in $\{-1, 0, 1\}$) holds since $Av_1 = (1, 0)^T$, $Av_2 = (-1, 0)^T$ and $Av_3 = (0, 1)^T$. Condition $(ii)$ of Theorem 3.2 also holds, as shown in the paragraph below.

**Verification of condition $(ii)$ of Theorem 3.2.** We first consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where

$$\begin{aligned}
&\partial_1(K_A + x) \cap \mathcal{X} \\
=\ & \{w \in \mathbb{Z}_+^4 \mid x_1 = w_1, x_2 \leq w_2\} \cap \mathcal{X} \\
=\ & \{w \in \mathbb{Z}_+^4 \mid x_1 = w_1, x_2 \leq w_2, x_1 + x_2 + x_4 = w_1 + w_2 + w_4 = \mathrm{S_{tot}}, x_3 + x_4 = w_3 + w_4 = \mathrm{E_{tot}}\} \\
=\ & \{w \in \mathbb{Z}_+^4 \mid x_1 = w_1, x_2 \leq w_2, x_3 \leq w_3, x_4 \geq w_4, w_1 + w_2 + w_4 = \mathrm{S_{tot}}, w_3 + w_4 = \mathrm{E_{tot}}\} \\
=\ & \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \leq w_2, x_3 \leq w_3, x_4 \geq w_4\}.
\end{aligned}$$

Since $\langle A_{1\bullet}, v_1 \rangle = 1$, $\langle A_{1\bullet}, v_2 \rangle = -1$, we need to check that $\Upsilon_1(x) \leq \check{\Upsilon}_1(y)$ and $\Upsilon_2(x) \geq \check{\Upsilon}_2(y)$. The first inequality holds because $y \in \partial_1(K_A + x) \cap \mathcal{X}$ implies $x_1 = y_1$ and $x_3 \leq y_3$ so that $\Upsilon_1(x) = \kappa_1 x_1 x_3 \leq \kappa_1 y_1 y_3 = \check{\kappa}_1 y_1 y_3 = \check{\Upsilon}_1(y)$. The second inequality holds because $y \in \partial_1(K_A + x) \cap \mathcal{X}$ implies $x_4 \geq y_4$ so that $\Upsilon_2(x) = \kappa_2 x_4 \geq \kappa_2 y_4 = \check{\kappa}_2 y_4 = \check{\Upsilon}_2(y)$. Secondly, we consider $x \in \mathcal{X}$, $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 = w_2, x_3 \geq w_3, x_4 \leq w_4\}$. Then, since $\langle A_{2\bullet}, v_3 \rangle = 1$, we need to check that $\Upsilon_3(x) \leq \check{\Upsilon}_3(y)$. This holds because $y \in \partial_2(K_A + x) \cap \mathcal{X}$ implies $x_4 \leq y_4$ so that $\Upsilon_3(x) = \kappa_3 x_4 \leq \kappa_3 y_4 \leq \check{\kappa}_3 y_4 = \check{\Upsilon}_3(y)$.

Since all of the hypotheses of Theorem 3.2 hold, we can conclude that, for each $x^\circ, \check{x}^\circ \in \mathcal{X}$ with $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with $\Upsilon$ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and

$$\mathbb{P}\left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0\right] = 1.$$

Furthermore, applying Theorem 3.4 with the set $\Gamma = \{p\} = \{(0, \mathrm{S_{tot}}, \mathrm{E_{tot}}, 0)\}$, which is increasing in $\mathcal{X}$ with respect to $\preccurlyeq_A$, we see that the mean first passage time from $s$ to $p$, $\mathbb{E}_s[T_p]$, is a decreasing function of $\kappa_3$.

Because the Markov chain has one absorbing state, $p$, per stoichiometric compatibility class, the stationary distribution on a given stoichiometric compatibility class is trivial, and hence so too are its monotonicity properties.

### Example 4.2. Enzyme kinetics II

Let us consider an extension of the enzymatic kinetics model introduced in the previous example, in which the substrate S can enter and leave the system and the product can revert to the substrate. This is a simplified version of the enzymatic kinetics considered by Anderson et al. [2]. The chemical reaction system is depicted in Fig. 3(a). Now, for this case study, we first determine how the reaction rate constant $\kappa_5$ affects the stochastic behavior of the system and then we will study properties of the system with respect to initial conditions.
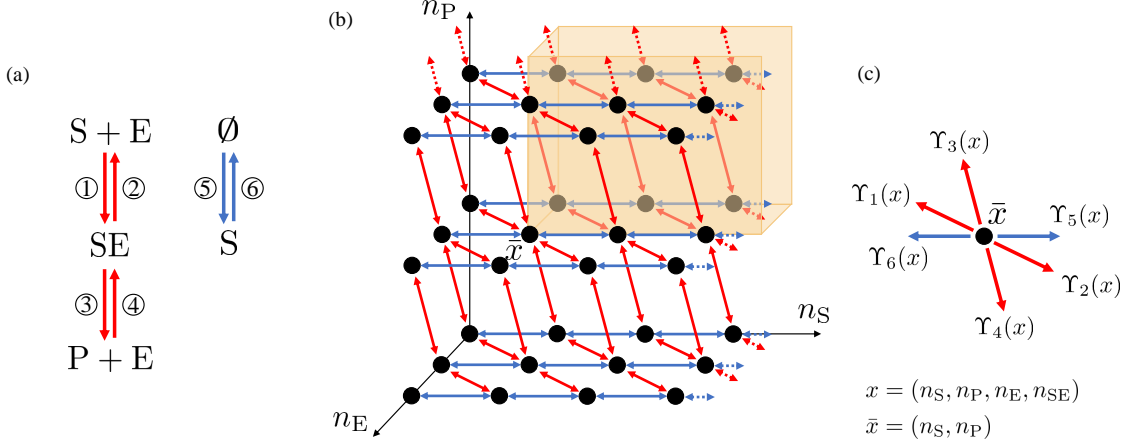
Figure 3: **Reaction model and corresponding Markov chain for enzymatic kinetics II example.** (a) Chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Projected Markov chain graph for one stoichiometric compatibility class with the conservation law $n_E + n_{SE} = E_{tot} = 2$. The projection takes a state $x = (n_S, n_P, n_E, n_{SE}) = (n_S, n_P, n_E, E_{tot} - n_E) \in \mathcal{X}$ to $\bar{x} = (n_S, n_P, n_E) \in \mathbb{Z}_+^3 : 0 \leq n_E \leq 2$. Here, we use black dots to represent the states, red double-ended arrows to represent transitions in both directions associated with the reactions represented by the red arrows in (a) and blue double-ended arrows to represent transitions in both directions associated with the reactions represented by the blue arrows in (a). We use dotted arrowed-lines to indicate that the pattern of Markov chain transitions extends to infinity. We use orange to highlight the projections of the region $K_A + x$ intersected with the stoichiometric compatibility class, where $A$ is defined in (4.4). (c) The projections of the directions of the possible transitions of the Markov chain within a stoichiometric compatibility class. The transition rates $\Upsilon_i(x)$, $i = 1, 2, 3, 4, 5, 6$, are defined in (4.3).

To this end, let us introduce the set of species $\mathscr{S} = \{S, P, E, SE\}$, and, similar to Example 4.1, we let $(n_S, n_P, n_E, n_{SE})$ be the state of the Markov chain that records the number of molecules of each species. The potential transitions of the Markov chain are in six possible directions, $v_j$ for $j = 1, ..., 6$, where $v_1 = -v_2 = (-1, 0, -1, 1)^T$, $v_3 = -v_4 = (0, 1, 1, -1)^T$, and $v_5 = -v_6 = (1, 0, 0, 0)^T$ (see SI-Section S.2.1 for the derivation of the $v_j$, $j = 1, ..., 6$). Since there is one linearly independent conservation law in this chemical reaction system: $n_E + n_{SE} = E_{tot}$, each stoichiometric compatibility class is contained in a three-dimensional affine subspace of four dimensional space, denoted as $z + \mathcal{L}$, where $z = (0, 0, E_{tot}, 0)$ and $\mathcal{L} := \text{span}\{v_1, v_3, v_5\}$, with fixed integer $E_{tot} > 0$. Then, we can choose the state space of the Markov chain to be $\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}_+^4 | x_3 + x_4 = E_{tot}\}$. Furthermore, given a state $x = (x_1, x_2, x_3, x_4) \in \mathcal{X}$, following mass-action kinetics, the associated infinitesimal transition rates are given by

$$
\begin{aligned}
\Upsilon_1(x) &= \kappa_1 x_1 x_3, \quad \Upsilon_2(x) = \kappa_2 x_4, \quad \Upsilon_3(x) = \kappa_3 x_4, \\
\Upsilon_4(x) &= \kappa_4 x_2 x_3, \quad \Upsilon_5(x) = \kappa_5, \quad \Upsilon_6(x) = \kappa_6 x_1,
\end{aligned}
\tag{4.3}
$$

for $\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5, \kappa_6 > 0$. As in Example 4.1, we apply our theory directly to our four dimensional Markov chain, but, for the purpose of illustration, Fig. 3(b) shows the three dimensional projection of the Markov chain graph for one stoichiometric compatibility class. Now, for the first analysis (determining how $\kappa_5$ affects the stochastic behavior of the system),

16

we verify that the assumptions of Theorems 3.2 and 3.5 hold and use them to determine how $\kappa_5$ affects the stationary distribution.

To this end, define the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{4.4}$$

and consider the preorder $x \preccurlyeq_A y$, defined by $A(y - x) \geq 0$. For $x \in \mathcal{X}$, $K_A + x = \{w \in \mathbb{R}^4 \mid x \preccurlyeq_A w\}$. Furthermore, let us consider the infinitesimal transition rates $\check{\Upsilon}_1(x)$, $\check{\Upsilon}_2(x)$, $\check{\Upsilon}_3(x)$, $\check{\Upsilon}_4(x)$, $\check{\Upsilon}_5(x)$ and $\check{\Upsilon}_6(x)$ defined as for $\Upsilon_1(x)$, $\Upsilon_2(x)$, $\Upsilon_3(x)$, $\Upsilon_4(x)$, $\Upsilon_5(x)$ and $\Upsilon_6(x)$, but with $\check{\kappa}_i$ in place of $\kappa_i$, where $\check{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4, 6$, and $\check{\kappa}_5 \geq \kappa_5$. Given that $Av_1 = (-1, 0, 1)^T$, $Av_2 = (1, 0, -1)^T$, $Av_3 = (0, 1, -1)^T$, $Av_4 = (0, -1, 1)^T$, $Av_5 = (1, 0, 0)^T$ and $Av_6 = (-1, 0, 0)^T$, we have that condition $(i)$ of Theorem 3.2 holds. Condition $(ii)$ of that theorem also holds, as shown in the next paragraph.

**Verification of condition $(ii)$ of Theorem 3.2.** First consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \leq w_2, x_3 \geq w_3, x_4 \leq w_4\}$. Since $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_5 \rangle = 1$ and $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_6 \rangle = -1$, we need to check that $\Upsilon_1(x) \geq \check{\Upsilon}_1(y), \Upsilon_6(x) \geq \check{\Upsilon}_6(y), \Upsilon_2(x) \leq \check{\Upsilon}_2(y)$, and $\Upsilon_5(x) \leq \check{\Upsilon}_5(y)$. Given that $y \in \partial_1(K_A + x) \cap \mathcal{X}$, the first inequality holds because $\Upsilon_1(x) = \kappa_1 x_1 x_3 \geq \kappa_1 y_1 y_3 = \check{\kappa}_1 y_1 y_3 = \check{\Upsilon}_1(y)$, the second inequality holds because $\Upsilon_6(x) = \kappa_6 x_1 = \kappa_6 y_1 = \check{\kappa}_6 y_1 = \check{\Upsilon}_6(y)$, the third inequality holds because $\Upsilon_2(x) = \kappa_2 x_4 \leq \kappa_2 y_4 = \check{\kappa}_2 y_4 = \check{\Upsilon}_2(y)$, and the fourth inequality holds because $\Upsilon_5(x) = \kappa_5 \leq \check{\kappa}_5 = \check{\Upsilon}_5(y)$. Secondly, we consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \leq w_1, x_2 = w_2, x_3 \geq w_3, x_4 \leq w_4\}$. Given that $\langle A_{3\bullet}, v_3 \rangle = 1$ and $\langle A_{3\bullet}, v_4 \rangle = -1$, we need to check that $\Upsilon_4(x) \geq \check{\Upsilon}_4(y)$ and $\Upsilon_3(x) \leq \check{\Upsilon}_3(y)$. The first inequality holds because $\Upsilon_4(x) = \kappa_4 x_2 x_3 \geq \kappa_4 y_2 y_3 = \check{\kappa}_4 y_2 y_3 = \check{\Upsilon}_4(y)$ and the second inequality holds because $\Upsilon_3(x) = \kappa_3 x_4 \leq \kappa_3 y_4 = \check{\kappa}_3 y_4 = \check{\Upsilon}_3(y)$. Finally, consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \leq w_1, x_2 \leq w_2, x_3 = w_3, x_4 = w_4\}$. Since $\langle A_{3\bullet}, v_1 \rangle = \langle A_{3\bullet}, v_4 \rangle = 1$ and $\langle A_{3\bullet}, v_2 \rangle = \langle A_{3\bullet}, v_3 \rangle = -1$, we need to check that $\Upsilon_2(x) \geq \check{\Upsilon}_2(y), \Upsilon_3(x) \geq \check{\Upsilon}_3(y), \Upsilon_1(x) \leq \check{\Upsilon}_1(y)$, and $\Upsilon_4(x) \leq \check{\Upsilon}_4(y)$. Indeed, we have that $\Upsilon_2(x) = \kappa_2 x_4 = \kappa_2 y_4 = \check{\kappa}_2 y_4 = \check{\Upsilon}_2(y), \Upsilon_3(x) = \kappa_3 x_4 = \kappa_3 y_4 = \check{\kappa}_3 y_4 = \check{\Upsilon}_3(y), \Upsilon_1(x) = \kappa_1 x_1 x_3 \leq \kappa_1 y_1 y_3 = \check{\kappa}_1 y_1 y_3 = \check{\Upsilon}_1(y)$, and $\Upsilon_4(x) = \kappa_4 x_2 x_3 \leq \kappa_4 y_2 y_3 = \check{\kappa}_4 y_2 y_3 = \check{\Upsilon}_4(y)$.

Thus, all of the hypothesis of Theorem 3.2 are verified, and so, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ satisfying $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with $\Upsilon$ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and $\mathbb{P}\left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0\right] = 1$. The Markov chains $X, \check{X}$ are irreducible and positive recurrent (see SI - Section S.1.1). Furthermore, for the increasing set in $\mathcal{X}$ with respect to $\preccurlyeq_A$ defined as $\Gamma(x) = \{w \in \mathcal{X} \mid x_1 \leq w_1, x_2 \leq w_2, x_3 \geq w_3, x_4 \leq w_4\}$, we can apply Theorem 3.5 and obtain that $\sum_{w \in \Gamma(x)} \pi_w \leq \sum_{w \in \Gamma(x)} \check{\pi}_w$. Loosely speaking, this means that increasing $\kappa_5$ causes the stationary distribution $\pi(x)$ to shift mass towards states characterized by lower $x_3$ and higher $x_1, x_2$ and $x_4$.

For this specific case, in which we have a stochastic chemical reaction network associated with a complex balanced dynamical system, an explicit expression for the stationary distribution can be obtained by applying Theorem 4.1 in Anderson et al. [2]. Analysis of this formula would provide results in agreement with the ones obtained by applying the theoretical tools developed in this paper. Specifically, $\pi_x$ can be written as a product of

two Poisson distributions and a binomial distribution, i.e.,

$$\pi_x = \left(e^{-c_1} \frac{c_1^{x_1}}{x_1!}\right) \left(e^{-c_2} \frac{c_2^{x_2}}{x_2!}\right) \left(\text{E}_{\text{tot}}! \frac{c_3^{x_3}}{x_3!} \frac{c_4^{x_4}}{x_4!}\right), \quad x \in \mathcal{X}, \tag{4.5}$$

in which $(c_1, c_2, c_3, c_4)$ represents the complex balanced equilibrium for the deterministic model, where

$$c_1 = \frac{\kappa_5}{\kappa_6}, \quad c_2 = \frac{\kappa_1 \kappa_3 \kappa_5}{\kappa_2 \kappa_4 \kappa_6}, \quad c_3 = \frac{1}{1 + \frac{\kappa_1 \kappa_5}{\kappa_2 \kappa_6}}, \text{ and } c_4 = \frac{\frac{\kappa_1 \kappa_5}{\kappa_2 \kappa_6}}{1 + \frac{\kappa_1 \kappa_5}{\kappa_2 \kappa_6}}. \tag{4.6}$$

In most cases, it is not possible to derive an analytical formula for the stationary distribution, but our theorems can still be applied and then monotonicity properties for $\pi$ can still be determined even without an explicit expression for $\pi$. For instance, in the context of the above example, if the infinitesimal transition rates $\Upsilon_i$ do not follow mass-action kinetics, the deficiency zero theorem and Theorem 4.1 in Anderson et al. [2] do not apply. Nevertheless, our theory can still be easily applied to study monotonicity properties for sample paths and stationary distributions.

As pointed out in Remark 3.4, we can also exploit our theoretical tools to determine monotonicity properties of the system with respect to the initial conditions. For this, suppose that $\breve{\kappa}_i = \kappa_i$ for $i = 1, 2, 3, 4, 5, 6$. Then, by the analysis above, Theorem 3.2 holds and yields monotonically (with preorder induced by the matrix $A$) with respect to the initial conditions.

**Example 4.3. A network topology arising in Braess' paradox** A natural question in synthetic biology may involve the prediction of whether an engineered biological circuit with additional reactions will lead to the desired effect of accelerating the process or unexpected behaviors. Now, we consider an example inspired by Braess' paradox, which arises from transportation networks, where adding one or more roads to a road network can slow down overall traffic flow through the network (see Braess [7] and see also a related state-dependent queuing network model in Calvert et al. [9]). A simple network of this type is one where there are two routes to get from the start to the final destination, and adding a linkage road between the routes can in some cases increase travel times. Fig. 4(a) shows a reaction network analogue of the Braess' network topology. Of course, our chemical reaction network is a little different from a road network since there is no congestion nor competition between molecules and pathways are chosen randomly with certain probabilities instead of routing decisions being based on the number of cars on the routes. Nevertheless, the example considered here is interesting because adding a reaction to cross-link two pathways might intuitively be interpreted as a detour and be expected to increase the time to the final destination, while this is sometimes not the case in this example.

The chemical reaction system is depicted in Fig. 4(a), which involves four species $\mathscr{S} = \{\text{S}_1, \text{S}_2, \text{S}_3, \text{S}_4\}$. The state of the Markov chain is $(n_{\text{S}_1}, n_{\text{S}_2}, n_{\text{S}_3}, n_{\text{S}_4})$ where $n_{\text{S}_i}$ is the number of copies of $\text{S}_i$ for $i = 1, 2, 3, 4$. The potential transitions of the Markov chain are in five possible directions, $v_j$, $j = 1, ..., 5$, where $v_1 = (-1, 1, 0, 0)^T$, $v_2 = (0, -1, 0, 1)^T$, $v_3 = (-1, 0, 1, 0)^T$, $v_4 = (0, 0, -1, 1)^T$ and $v_5 = (0, -1, 1, 0)^T$ (see SI-Section S.2.2 for the derivation of the $v_j$, $j = 1, ..., 5$). Fixing an integer $\text{S}_{\text{tot}} > 0$, the associated stoichiometric compatibility class is $z + \mathcal{L}$ with $z = (\text{S}_{\text{tot}}, 0, 0, 0)$ and $\mathcal{L} := \text{span}\{v_1, v_2, v_3, v_4, v_5\}$. The set $z + \mathcal{L}$ is a three-dimensional affine subspace of four dimensional space. We choose the state
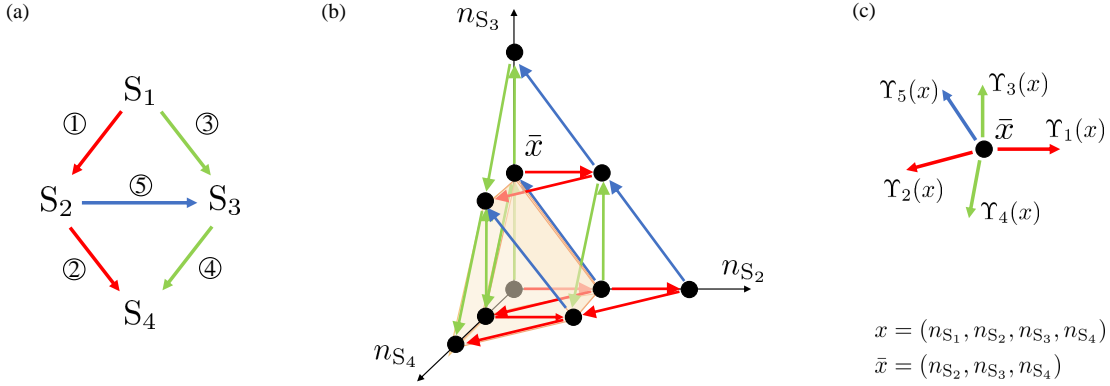
Figure 4: **Circuit inspired by Braess' paradox and corresponding Markov chain.** (a) Chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Projected Markov chain graph for one stoichiometric compatibility class with the conservation law $n_{S_1} + n_{S_2} + n_{S_3} + n_{S_4} = S_{tot} = 2$. The projection takes a state $x = (n_{S_1}, n_{S_2}, n_{S_3}, n_{S_4}) = (S_{tot} - n_{S_2} - n_{S_3} - n_{S_4}, n_{S_2}, n_{S_3}, n_{S_4}) \in \mathcal{X}$ to $\bar{x} = (n_{S_2}, n_{S_3}, n_{S_4})$. Here, we use black dots to represent the states and red (blue, green) arrows to represent transitions in directions associated with the reactions represented by the red (blue, green) arrows in (a). We use orange to highlight the projection of the region $K_A + x$ intersected with the stoichiometric compatibility class, where $A$ is defined in (4.3). (c) The projections of the directions of the possible transitions of the Markov chain within a stoichiometric compatibility class. The transition rates $\Upsilon_i(x)$, $i = 1, 2, 3, 4, 5$, are given in (4.7).

space of our Markov chain to be $\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}_+^4 | x_1 + x_2 + x_3 + x_4 = S_{tot}\}$. The constraint introduced in the last expression for $\mathcal{X}$ follows from the conservation law in this chemical reaction system, that is $n_{S_1} + n_{S_2} + n_{S_3} + n_{S_4} = S_{tot}$. Given a generic state $x = (x_1, x_2, x_3, x_4)$, following mass-action kinetics, the infinitesimal transition rates are

$$\Upsilon_1(x) = \kappa_1 x_1, \quad \Upsilon_2(x) = \kappa_2 x_2, \quad \Upsilon_3(x) = \kappa_3 x_1, \quad \Upsilon_4(x) = \kappa_4 x_3, \quad \Upsilon_5(x) = \kappa_5 x_2. \quad (4.7)$$

For the purpose of illustration, Fig. 4(b) shows the three dimensional projection of the Markov chain graph for one stoichiometric compatibility class.

A natural question is how the time $T_{(0,0,0,S_{tot})}$ to reach the state $(0, 0, 0, S_{tot})$ from $(S_{tot}, 0, 0, 0)$ depends on the rate constants $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ and $\kappa_5$. For this, we use Theorem 3.4. Let

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

The matrix $A$ here defines a preorder that is not a partial order of $\mathcal{X}$. For $x \in \mathcal{X}$, consider infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x), \check{\Upsilon}_3(x), \check{\Upsilon}_4(x)$ and $\check{\Upsilon}_5(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x)$ and $\Upsilon_5(x)$, but with $\check{\kappa}_i$ in place of $\kappa_i$ where $\check{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4$, and $\check{\kappa}_5 \neq \kappa_5$. Suppose that $\kappa_2 = \kappa_4$. Now, let us verify that the assumptions of Theorem 3.3 hold. Condition $(i)$ holds since $Av_1 = (1, -1)^T$, $Av_2 = (0, 1)^T$, $Av_3 = (1, -1)^T$, $Av_4 = (0, 1)^T$ and $Av_5 = (0, 0)^T$. Condition $(ii)$ of Theorem 3.3 also holds, as shown in the paragraph below.

19

**Verification of condition** (*ii*) **of Theorem 3.3.** Let $x \in \mathcal{X}$, and first consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \,|\, x_1 = w_1, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $Av_2 = Av_4$, $Av_1 = Av_3$, and $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \leq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$. Since $y \in \partial_1(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 = \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = \kappa_3 x_1 = \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$, and so the desired inequality holds with equality. Secondly, consider $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \,|\, x_1 \geq w_1, x_2 + x_3 = w_2 + w_3, x_4 \leq w_4\}$. Given that $Av_2 = Av_4$, $Av_1 = Av_3$, and $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = -1$ and $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_4 \rangle = 1$, we need to check that $\Upsilon_2(x) + \Upsilon_4(x) \leq \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$ and $\Upsilon_1(x) + \Upsilon_3(x) \geq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$. For $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X}$, we have that $\Upsilon_2(x) + \Upsilon_4(x) = \kappa_2 x_2 + \kappa_4 x_3 = \kappa_2(x_2 + x_3) \leq \kappa_2(y_2 + y_3) = \check{\kappa}_2(y_2 + y_3) = \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$ and $\Upsilon_1(x) = \kappa_1 x_1 \geq \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$, $\Upsilon_3(x) = \kappa_3 x_1 \geq \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$.

Thus, all hypotheses of Theorem 3.3 hold, and so for every $x^\circ, \check{x}^\circ \in \mathcal{X}$ where $x^\circ \preccurlyeq_A \check{x}^\circ$ there there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with $\Upsilon$ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and $\mathbb{P}\left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0\right] = 1$. Let $\Gamma = \{(0, 0, 0, \mathrm{S_{tot}})\}$. This is an increasing set in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$. Let $T_{(0,0,0,\mathrm{S_{tot}})}$, respectively $\check{T}_{(0,0,0,\mathrm{S_{tot}})}$ be the first time that the Markov chain $X$, respectively $\check{X}$, reaches the set $\Gamma$. Then, by Theorem 3.4, if $X(0) = \check{X}(0) = (\mathrm{S_{tot}}, 0, 0, 0)$, we have that $\check{T}_{(0,0,0,\mathrm{S_{tot}})} \preccurlyeq_{st} T_{(0,0,0,\mathrm{S_{tot}})}$. By interchanging $\check{\Upsilon}_5$ and $\kappa_5$, we can conclude that $\check{T}_{(0,0,0,\mathrm{S_{tot}})}$ and $T_{(0,0,0,\mathrm{S_{tot}})}$ are stochastically equivalent (equal in distribution). It follows that the mean first passage time from $(\mathrm{S_{tot}}, 0, 0, 0)$ to $(0, 0, 0, \mathrm{S_{tot}})$ is insensitive to $\kappa_5$ when $\kappa_2 = \kappa_4$. This is naively counter-intuitive: since the fifth reaction re-routes some samples to another state where the last reaction has the same rate constant as the final reaction without re-routing, it should take a longer expected time since re-routing also takes some time. However, in reality, the presence of the fifth reaction also fastens the rate to transition from $\mathrm{S}_2$, and this balances the time of re-routing. Most importantly, our theorem is able to capture this result without explicitly calculating the mean first passage time and allows us to reach the conclusion easily. We expect that in more complex situations, our method will be a valuable tool to establish monotonicity and insensitivity results.

Given that the Markov chain has one absorbing state per stoichiometric compatibility class, the stationary distribution for a given stoichiometric compatibility class is trivial, and hence so too are its monotonicity properties.

Theorem S.2 allows us to conclude further interesting properties for this network. Using two other $A$ matrices (see SI - Section S.3.2), we can conclude that adding reaction ⑤ (changing from $\kappa_5 = 0$ to $\kappa_5 > 0$) causes the mean first passage time from $(\mathrm{S_{tot}}, 0, 0, 0)$ to $(0, 0, 0, \mathrm{S_{tot}})$ to increase if $\kappa_2 > \kappa_4$ or to decrease if $\kappa_2 < \kappa_4$. More explicitly, this shows that there can be opposing effects on the mean first passage time with different choices of $\kappa_2$ and $\kappa_4$ when reaction ⑤ is added.

**Example 4.4. Epigenetic regulation by chromation modifications**

Epigenetic regulation is the modification of the DNA structure, due to chromatin modifications, that determines if a gene is active or repressed. There are several chromatin modifications that can affect the DNA structure. Here, we will focus only on histone modifications. More precisely, we consider a ubiquitous model for a histone modification circuit (see Dodd et al. [11] and Bruno et al. [8]). The species considered are nucleosomes that are unmodified (D), modified with repressive modifications ($\mathrm{D^R}$), and modified with activating
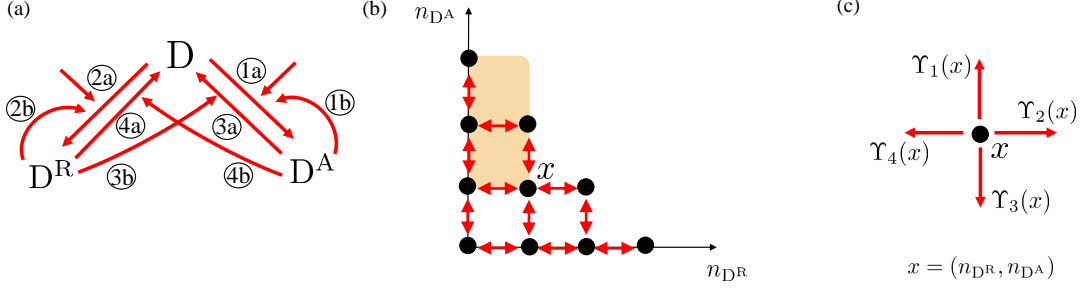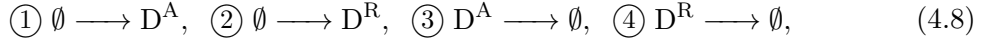
20

Figure 5: **Histone modification circuit and corresponding Markov chain.** (a) Original chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Markov chain graph associated with the reduced chemical reaction system. Here, we consider $D_{tot} = 3$, we use black dots to represent the states and red double-ended arrows to represent transitions in both directions. We use orange to highlight the region $(K_A + x) \cap \mathcal{X}$, with $A$ defined in (4.11). (c) Direction of the possible transitions of the Markov chains, whose rates are given in equation (4.10).

modifications ($D^A$), and, in terms of molecular interactions, each histone modification autocatalyzes itself and promotes the erasure of the other one. The chemical reaction system considered is depicted in Fig. 5(a). The amount of each species is represented by $n_D$, $n_{D^R}$ and $n_{D^A}$, respectively, and their sum is conserved, that is $n_D + n_{D^R} + n_{D^A} = D_{tot}$, with $D_{tot}$ representing the total number of nucleosomes within the gene.

By fixing an integer $D_{tot} > 0$, we fix one stoichiometric compatibility class. The projected process $(X_1, X_2)(\cdot) = (n_{D^R}, n_{D^A})$ is still a continuous-time Markov chain, and in this example we choose to apply our theory to this reduced system. This is the same as studying the reduced chemical reaction system defined as follows:

$$ \text{①} \; \emptyset \longrightarrow D^A, \quad \text{②} \; \emptyset \longrightarrow D^R, \quad \text{③} \; D^A \longrightarrow \emptyset, \quad \text{④} \; D^R \longrightarrow \emptyset, \qquad (4.8) $$

with two species $\mathscr{S} = \{D^R, D^A\}$ and four reactions $\mathscr{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+)\}$, where $v_1^- = v_2^- = v_3^+ = v_4^+ = (0,0)^T$, $v_2^+ = v_4^- = (1,0)^T$, $v_1^+ = v_3^- = (0,1)^T$, and with associated propensity functions of non mass-action type defined as follows:

$$ \Lambda_{(v_1^-, v_1^+)}(x) = (D_{tot} - (x_1 + x_2))(\kappa_{1a} + \kappa_{1b} x_2), $$
$$ \Lambda_{(v_2^-, v_2^+)}(x) = (D_{tot} - (x_1 + x_2))(\kappa_{2a} + \kappa_{2b} x_1), \qquad (4.9) $$
$$ \Lambda_{(v_3^-, v_3^+)}(x) = x_2(\kappa_{3a} + x_1 \kappa_{3b}), \quad \Lambda_{(v_4^-, v_4^+)}(x) = x_1 \mu (c\kappa_{3a} + x_2 \kappa_{3b}), $$

in which $\kappa_{1a}$, $\kappa_{1b}$, $\kappa_{3a}$, $\kappa_{3b}$, $\kappa_{2a}$, $\kappa_{2b}$, $\kappa_{4a} = \mu c \kappa_{3a}$, $\kappa_{4b} = \mu \kappa_{3b}$ are the rate constants that go with each of the reactions shown in Fig. 5(a), respectively.

The state space for the Markov chain is $\mathcal{X} = \{(x_1, x_2) \in \mathbb{Z}_+^2 \mid x_1 + x_2 \leq D_{tot}\}$. Given a generic state $x = (x_1, x_2) \in \mathcal{X}$, the potential transitions of the Markov chain are in four possible directions $v_j = v_j^+ - v_j^-$, $j = 1, 2, 3, 4$, that can be written as $v_1 = (0,1)^T$, $v_2 = (1,0)^T$, $v_3 = (0,-1)^T$ and $v_4 = (-1,0)^T$, with associated infinitesimal transition rates

$$ \Upsilon_1(x) = \Lambda_{(v_1^-, v_1^+)}(x), \quad \Upsilon_2(x) = \Lambda_{(v_2^-, v_2^+)}(x), \quad \Upsilon_3(x) = \Lambda_{(v_3^-, v_3^+)}(x), \quad \Upsilon_4(x) = \Lambda_{(v_4^-, v_4^+)}(x). \tag{4.10} $$

We are interested in determining how the asymmetry of the system, represented by the parameter $\mu$ affects the stochastic behavior of the system. In particular, we will focus

on studying the stationary distribution and the *time to memory loss* of the active and repressed state, defined as the mean first passage time to reach the fully repressed state ($r = (n_{\mathrm{DR}}, n_{\mathrm{DA}}) = (\mathrm{D}_{\mathrm{tot}}, 0)$), starting from the fully active state ($a = (n_{\mathrm{DR}}, n_{\mathrm{DA}}) = (0, \mathrm{D}_{\mathrm{tot}})$), and vice versa (i.e., $h_{a,r} = \mathbb{E}_a[T_r]$ and $h_{r,a} = \mathbb{E}_r[T_a]$). To this end, we first verify that we can apply Theorem 3.2.

Let

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{4.11}$$

For $x \in \mathcal{X}$, $K_A + x = \{w \in \mathbb{R}^2 | x \preccurlyeq_A w\}$ and $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} | x \preccurlyeq_A w\}$. See Fig. 5(b) for an example of $\mathcal{X}$ and $(K_A + x) \cap \mathcal{X}$ for $\mathrm{D}_{\mathrm{tot}} = 3$. We introduce infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x), \check{\Upsilon}_3(x)$ and $\check{\Upsilon}_4(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x)$ and $\Upsilon_4(x)$, with all the parameters having the same values except that $\mu$ is replaced by $\check{\mu}$, where $\check{\mu} \geq \mu$. Since $Av_1 = (0,1)^T$, $Av_2 = (-1,0)^T$, $Av_3 = (0,-1)^T$ and $Av_4 = (1,0)^T$, we have that condition $(i)$ of Theorem 3.2 holds. Condition $(ii)$ also holds, as shown in the paragraph below.

**Verification of condition $(ii)$ of Theorem 3.2.** Consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A+x) \cap \mathcal{X}$, where $\partial_1(K_A+x) \cap \mathcal{X} = \{w \in \mathcal{X} | x_1 = w_1, x_2 \leq w_2\}$. Since $\langle A_{1\bullet}, v_4 \rangle = 1$ and $\langle A_{1\bullet}, v_2 \rangle = -1$, we must check that $\Upsilon_2(x) \geq \check{\Upsilon}_2(y)$ and $\Upsilon_4(x) \leq \check{\Upsilon}_4(y)$. Since $y \in \partial_1(K_A+x) \cap \mathcal{X}$ implies $x_1 = y_1$ and $x_2 \leq y_2$, we have $\Upsilon_2(x) = (\mathrm{D}_{\mathrm{tot}} - (x_1 + x_2))(\kappa_{2a} + \kappa_{2b}x_1) \geq (\mathrm{D}_{\mathrm{tot}} - (y_1 + y_2))(\kappa_{2a} + \kappa_{2b}y_1) = \check{\Upsilon}_2(y)$ and $\Upsilon_4(x) = x_1\mu(c\kappa_{3a} + x_2\kappa_{3b}) \leq y_1\mu(c\kappa_{3a} + y_2\kappa_{3b}) \leq y_1\check{\mu}(c\kappa_{3a} + y_2\kappa_{3b}) = \check{\Upsilon}_4(y)$, and so both inequalities hold. Similarly, for $x \in \mathcal{X}$ and $y \in \partial_2(K_A+x) \cap \mathcal{X} = \{w \in \mathcal{X} | x_1 \geq w_1, x_2 = w_2\}$, since $\langle A_{2\bullet}, v_1 \rangle = 1$ and $\langle A_{2\bullet}, v_3 \rangle = -1$, we need to check that $\Upsilon_1(x) \leq \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) \geq \check{\Upsilon}_3(y)$. Indeed, $\Upsilon_1(x) = (\mathrm{D}_{\mathrm{tot}} - (x_1 + x_2))(\kappa_{1a} + \kappa_{1b}x_2) \leq (\mathrm{D}_{\mathrm{tot}} - (y_1+y_2))(\kappa_{1a} + \kappa_{1b}y_2) = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = x_2(\kappa_{3a} + x_1\kappa_{3b}) \geq y_2(\kappa_{3a} + y_1\kappa_{3b}) = \check{\Upsilon}_3(y)$.

Since all of the hypotheses of Theorem 3.2 hold, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ satisfying $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with $\Upsilon$ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and $\mathbb{P}\left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0\right] = 1$.

We can also apply Theorem 3.5. The Markov chains $X$ and $\check{X}$ are irreducible and, having only finitely many states, are positive recurrent. Based on the order $\preccurlyeq_A$ we introduced, the fully active state $a = (0, \mathrm{D}_{\mathrm{tot}})$ is maximal in $\mathcal{X}$ and the fully repressed state $r = (\mathrm{D}_{\mathrm{tot}}, 0)$ is minimal in $\mathcal{X}$. Then, by Theorem 3.5, we can conclude that $\pi_a \leq \check{\pi}_a$ and $\pi_r \geq \check{\pi}_r$. This implies that increasing $\mu$ increases the probability of the system in steady-state to be in the active state $a$ to the detriment of the repressed state $r$ (and vice versa for decreasing $\mu$). We can also apply Theorem 3.4. Since $\{a\}$ is increasing and $\{r\}$ is decreasing, then by Theorem 3.4, $\check{h}_{r,a} = \mathbb{E}_r[\check{T}_a] \leq \mathbb{E}_r[T_a] = h_{r,a}$ and $h_{a,r} = \mathbb{E}_a[T_r] \leq \mathbb{E}_a[\check{T}_r] = \check{h}_{a,r}$. Since the only difference between the two systems was that $\mu \leq \check{\mu}$, these results imply that the time to memory loss of the active state increases for higher values of $\mu$, while the time to memory loss of the repressed state decreases for higher values of $\mu$.

### Example 4.5. Epigenetic regulation by chromatin modifications with positive TF-enabled autoregulation

Now, we consider the histone modification circuit considered in the previous example with an additional positive autoregulation loop. For this, we assume that a protein expressed by the gene of interest recruits writers for the activating histone modifications. Consequently, we introduce the gene product P as an additional species for our system and add the
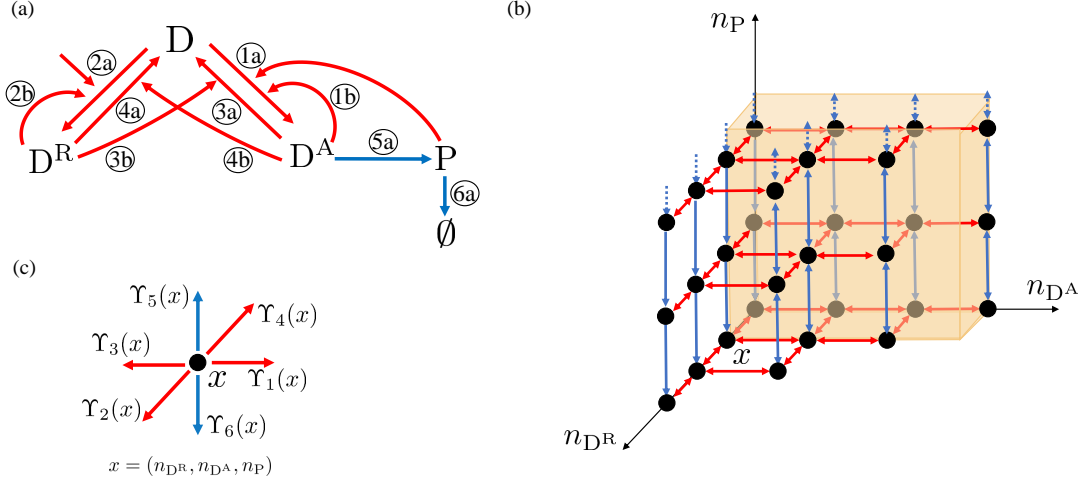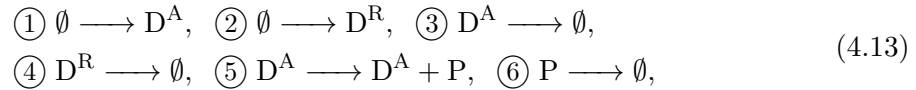
Figure 6: **Histone modification circuit with positive TF-enabled autoregulation and corresponding Markov chain.** (a) Original chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Markov chain graph. Here, we consider $D_{tot} = 3$, we use black dots to represent the states and red double-ended arrows to represent transitions in both directions associated with the reactions represented by the red arrows in (a). Similarly we use blue double-ended (single-ended) arrows to represent transitions in both directions (in one direction) associated with the reactions represented by the blue arrows in (a). We use blue dotted lines to show that, in the vertical direction, the Markov chain has countably infinitely many states, connected by transitions in both directions. Finally, we use orange to highlight the region $K_A + x$ intersected with the state space $\mathcal{X}$, with $A$ defined in (4.15). (c) Direction of the possible transitions of the Markov chain starting from a state $x$, whose rates are defined in equation (4.14).

following reactions to the ones shown in Fig. 5(a):

$$\text{⑤a } D^A \longrightarrow D^A + P, \quad \text{⑥a } P \longrightarrow \emptyset. \tag{4.12}$$
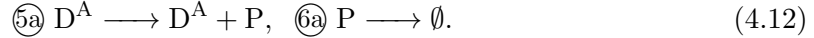
Furthermore, given the P-enabled autoregulation loop (Fig. 6(a)), let us consider the rate constant that goes with ①a in Fig. 5(a) as $\kappa_{1a} = \kappa_{1a}^0 + \kappa_{1a}^1 g(n_P)$, with $\kappa_{1a}^0$ and $\kappa_{1a}^1$ representing the rate constants that go with the $D^A$ basal *de-novo* establishment process and with the $D^A$ *de-novo* establishment process enhanced by $n_P$, respectively, and $g(n_P)$ representing a non-negative, bounded, monotonically increasing function of $n_P$ (see Bruno et al. [8], Section 3.4).

Here, we are interested in determining how the reaction rate constant $\kappa_{5a}$ affects the reactivation time of the gene. As before, we have the conservation law $n_D + n_{D^R} + n_{D^A} = D_{tot}$, with $D_{tot}$ representing the total number of nucleosomes within the gene, and by fixing $D_{tot} > 0$, we fix one stoichiometric compatibility class and the projected process $(X_1, X_2, X_3)(\cdot) = (n_{D^R}, n_{D^A}, n_P)$ is a continuous-time Markov chain. This is the same as studying the reduced chemical reaction system:

$$\begin{aligned} &\text{① } \emptyset \longrightarrow D^A, \quad \text{② } \emptyset \longrightarrow D^R, \quad \text{③ } D^A \longrightarrow \emptyset, \\ &\text{④ } D^R \longrightarrow \emptyset, \quad \text{⑤ } D^A \longrightarrow D^A + P, \quad \text{⑥ } P \longrightarrow \emptyset, \end{aligned} \tag{4.13}$$

with set of species $\mathscr{S} = \{D^R, D^A, P\}$, set of reactions $\mathscr{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+), (v_5^-, v_5^+), (v_6^-, v_6^+)\}$, where $v_1^- = v_2^- = v_3^+ = v_4^+ = v_6^+ = (0, 0, 0)^T$, $v_2^+ = v_4^- =$

23

$(1,0,0)^T$, $v_1^+ = v_3^- = v_5^- = (0,1,0)^T$, $v_5^+ = (0,1,1)^T$, $v_6^- = (0,0,1)^T$, and with associated propensity functions of non mass-action type defined as follows:

$$\Lambda_{(v_1^-,v_1^+)}(x) = (\mathrm{D_{tot}} - (x_1 + x_2))\left(\kappa_{1a}^0 + \kappa_{1a}^1 g(x_3) + \kappa_{1b}x_2\right),$$

$$\Lambda_{(v_2^-,v_2^+)}(x) = (\mathrm{D_{tot}} - (x_1 + x_2))(\kappa_{2a} + \kappa_{2b}x_1), \quad \Lambda_{(v_3^-,v_3^+)}(x) = x_2(\kappa_{3a} + x_1\kappa_{3b}),$$

$$\Lambda_{(v_4^-,v_4^+)}(x) = x_1\mu(c\kappa_{3a} + x_2\kappa_{3b}), \quad \Lambda_{(v_5^-,v_5^+)}(x) = \kappa_{5a}x_2, \quad \Lambda_{(v_6^-,v_6^+)}(x) = \kappa_{6a}x_3,$$

in which $\kappa_{5a}$ and $\kappa_{6a}$ are the rate constants that go with reactions ⑤ⓐ and ⑥ⓐ in (4.12), respectively, and all the other rate constants are defined as for (4.9).

The state space for the Markov chain is $\mathcal{X} = \{(x_1, x_2, x_3) \in \mathbb{Z}_+^3 \mid x_1 + x_2 \leq \mathrm{D_{tot}}\}$. Given a generic state $x = (x_1, x_2, x_3)$, the transitions of the Markov chain are in six possible directions $v_j = v_j^+ - v_j^-$, $j \in \{1, ..., 6\}$, that can be written as $v_1 = (0, 1, 0)^T$, $v_2 = (1, 0, 0)^T$, $v_3 = (0, -1, 0)^T$, $v_4 = (-1, 0, 0)^T$, $v_5 = (0, 0, 1)^T$, $v_6 = (0, 0, -1)^T$, with associated infinitesimal transition rates:

$$\Upsilon_1(x) = \Lambda_{(v_1^-,v_1^+)}(x), \quad \Upsilon_2(x) = \Lambda_{(v_2^-,v_2^+)}(x), \quad \Upsilon_3(x) = \Lambda_{(v_3^-,v_3^+)}(x),$$

$$\Upsilon_4(x) = \Lambda_{(v_4^-,v_4^+)}(x), \quad \Upsilon_5(x) = \Lambda_{(v_5^-,v_5^+)}(x), \quad \Upsilon_6(x) = \Lambda_{(v_6^-,v_6^+)}(x). \tag{4.14}$$

As mentioned before, we are interested in determining how the protein production rate $\kappa_{5a}$ affects the reactivation time of the gene, defined as $h_{r,\Theta} = \mathbb{E}_r[T_\Theta]$, where $r = (\mathrm{D_{tot}}, 0, 0)$ and $\Theta = \{w \in \mathcal{X} \mid w = (0, \mathrm{D_{tot}}, i), i \in \mathbb{Z}_+\}$ corresponds to the set of states characterized by the fully active state $n_{\mathrm{D^A}} = \mathrm{D_{tot}}$. We first check that the assumptions of Theorem 3.2 hold. Let

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.15}$$

For $x \in \mathcal{X}$, $x \preccurlyeq_A y$ and the set $K_A + x = \{y \in \mathbb{R}^3 \mid x \preccurlyeq_A y\}$. For our example, the region $(K_A + x) \cap \mathcal{X}$ is depicted in orange in Fig. 6(b). We introduce infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x), \check{\Upsilon}_3(x), \check{\Upsilon}_4(x), \check{\Upsilon}_5(x)$ and $\check{\Upsilon}_6(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x), \Upsilon_5(x)$ and $\Upsilon_6(x)$, with all the parameters having the same values except that $\kappa_{5a}$ is replaced by $\check{\kappa}_{5a} > \kappa_{5a}$. Condition $(i)$ of Theorem 3.2 holds since $Av_1 = (0, 1, 0)^T$, $Av_2 = (-1, 0, 0)^T$, $Av_3 = (0, -1, 0)^T$, $Av_4 = (1, 0, 0)^T$, $Av_5 = (0, 0, 1)^T$, $Av_6 = (0, 0, -1)^T$. Condition $(ii)$ also holds, as shown in the paragraph below.

**Verification of condition $(ii)$ of Theorem 3.2.** First consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \leq w_2, x_3 \leq w_3\}$. Since $\langle A_{1\bullet}, v_4 \rangle = 1$ and $\langle A_{1\bullet}, v_2 \rangle = -1$, we need to check that $\Upsilon_4(x) \leq \check{\Upsilon}_4(y)$ and $\Upsilon_2(x) \geq \check{\Upsilon}_2(y)$. Since $x_1 = y_1, x_2 \leq y_2, x_3 \leq y_3$, we have that $\Upsilon_4(x) = x_1\mu(c\kappa_{3a} + x_2\kappa_{3b}) \leq y_1\mu(c\kappa_{3a} + y_2\kappa_{3b}) = \check{\Upsilon}_4(y)$ and $\Upsilon_2(x) = (\mathrm{D_{tot}} - (x_1 + x_2))(\kappa_{2a} + \kappa_{2b}x_1) \geq (\mathrm{D_{tot}} - (y_1 + y_2))(\kappa_{2a} + \kappa_{2b}y_1) = \check{\Upsilon}_2(y)$. Secondly, consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 = w_2, x_3 \leq w_3\}$. Since $\langle A_{2\bullet}, v_1 \rangle = 1$ and $\langle A_{2\bullet}, v_3 \rangle = -1$, we need to check that $\Upsilon_1(x) \leq \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) \geq \check{\Upsilon}_3(y)$. Since $x_1 \geq y_1, x_2 = y_2, x_3 \leq y_3$, we have $\Upsilon_1(x) = (\mathrm{D_{tot}} - (x_1 + x_2))\left(\kappa_{1a}^0 + \kappa_{1a}^1 g(x_3) + \kappa_{1b}x_2\right) \leq (\mathrm{D_{tot}} - (y_1 + y_2))\left(\kappa_{1a}^0 + \kappa_{1a}^1 g(y_3) + \kappa_{1b}y_2\right) = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = x_2(\kappa_{3a} + x_1\kappa_{3b}) \geq y_2(\kappa_{3a} + y_1\kappa_{3b}) = \check{\Upsilon}_3(y)$. Finally, consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 \leq w_2, x_3 = w_3\}$. Since $\langle A_{3\bullet}, v_5 \rangle = 1$ and $\langle A_{3\bullet}, v_6 \rangle = -1$, we must check that $\Upsilon_5(x) \leq \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) \geq \check{\Upsilon}_6(y)$. Since $x_1 \geq y_1, x_2 \leq y_2, x_3 = y_3$, we obtain $\Upsilon_5(x) = \kappa_{5a}x_2 \leq \kappa_{5a}y_2 \leq \check{\kappa}_{5a}y_2 = \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) = \kappa_{6a}x_3 = \kappa_{6a}y_3 = \check{\Upsilon}_6(y)$.

Since all the hypotheses of Theorem 3.2 hold, for each $x^\circ, \breve{x}^\circ \in \mathcal{X}$ satisfying $x^\circ \preccurlyeq_A \breve{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), \, t \geq 0\}$ and $\breve{X} = \{\breve{X}(t), \, t \geq 0\}$ associated with $\Upsilon$ and $\breve{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\breve{X}(0) = \breve{x}^\circ$ and $\mathbb{P}\left[X(t) \preccurlyeq_A \breve{X}(t) \text{ for every } t \geq 0\right] = 1$.

Furthermore, since the hypotheses of Theorem 3.2 hold, we can also apply Theorem 3.4. Specifically, for $r = (\mathrm{D}_{\mathrm{tot}}, 0, 0)$ and $\Theta = \{y \in \mathcal{X} | y = (0, \mathrm{D}_{\mathrm{tot}}, i), i \in \mathbb{Z}_+\}$, since $\Theta$ is an increasing set in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$, then $h_{r,\Theta} \geq \breve{h}_{r,\Theta}$. This implies that, assuming that the only difference between the two systems is in the value of the protein production rate parameter, $\kappa_{5a}$, higher protein production rates reduce the mean reaction time for the gene.

# 5 Proofs of the Main Results

## 5.1 Proof of Theorem 3.1.

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\breve{\Upsilon} = (\breve{\Upsilon}_1, \ldots, \breve{\Upsilon}_n)$, such that (3.3) holds. Let $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$ and $\breve{Q} = (\breve{Q}_{x,y})_{x,y \in \mathcal{X}}$ denote the infinitesimal generators for the continuous-time Markov chains associated with $\Upsilon$ and $\breve{\Upsilon}$, respectively. In the following, let $A \in \mathbb{R}^{m \times d}$ be a matrix with non-zero rows and consider the relation $\preccurlyeq_A$ as defined in Definition 3.1.

For the proof of Theorem 3.1, we first assume that

$$\sup_{x \in \mathcal{X}} \Upsilon_j(x) < \infty \quad \text{and} \quad \sup_{x \in \mathcal{X}} \breve{\Upsilon}_j(x) < \infty \qquad \text{for every } 1 \leq j \leq n. \tag{5.1}$$

This restriction will be relaxed later. Then, we define a constant $\lambda > 0$ and a pair of functions $\Phi_\lambda$ and $\breve{\Phi}_\lambda$, which will be key to our construction of the coupled processes $X$ and $\breve{X}$. Let $\lambda > 0$ such that:

$$\lambda > n \max \left\{ \sup_{x \in \mathcal{X}} \sum_{j=1}^n \Upsilon_j(x), \sup_{x \in \mathcal{X}} \sum_{j=1}^n \breve{\Upsilon}_j(x) \right\}. \tag{5.2}$$

Note that both $\frac{\Upsilon_j(x)}{\lambda}$ and $\frac{\breve{\Upsilon}_j(x)}{\lambda}$ are less than $\frac{1}{n}$ for every $x \in \mathcal{X}$ and $1 \leq j \leq n$. For $x \in \mathcal{X}$, consider the sets

$$I_j(x) := \left[ \frac{j-1}{n}, \frac{j-1}{n} + \frac{\Upsilon_j(x)}{\lambda} \right), \qquad 1 \leq j \leq n. \tag{5.3}$$

If $\Upsilon_j(x) = 0$, then $I_j(x)$ is the empty set. On the other hand, if $\Upsilon_j(x) > 0$, then $I_j(x)$ is an interval that is a strict subset of $[\frac{j-1}{n}, \frac{j}{n})$. Define the function $\Phi_\lambda(\cdot, \cdot) : \mathcal{X} \times [0,1] \longrightarrow \mathcal{X}$ by

$$\Phi_\lambda(x, u) := x + \sum_{j=1}^n v_j \mathbb{1}_{I_j(x)}(u), \qquad x \in \mathcal{X}, \ u \in [0,1]. \tag{5.4}$$

For $x \in \mathcal{X}$, the sets $I_1(x), \ldots, I_n(x)$ are mutually disjoint and so for any $u \in [0,1]$ either $\Phi_\lambda(x, u) = x$ or $\Phi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$. In the second case, this will happen

25

if and only if $u \in I_j(x)$ for the corresponding index $j$. The latter condition implies that $I_j(x) \neq \emptyset$, hence by (5.3), $\Upsilon_j(x) > 0$ and by (3.3), $x + v_j \in \mathcal{X}$.

This shows that $\Phi_\lambda(\cdot, \cdot)$ is well-defined as an $\mathcal{X}$-valued function. We define intervals $\breve{I}_j(x)$, $1 \leq j \leq n$, $x \in \mathcal{X}$ and a function $\breve{\Phi}_\lambda : \mathcal{X} \times [0,1] \longrightarrow \mathcal{X}$ in an analogous manner to that above, where $\breve{\Phi}_\lambda$ is defined as in (5.4), but with the intervals $I_j(x)$ replaced by $\breve{I}_j(x)$, where these are defined as in (5.3), but with $\Upsilon_j(x)$ replaced by $\breve{\Upsilon}_j(x)$.

**Lemma 5.1.** *Suppose that $x, y \in \mathcal{X}$ are such that $x \preccurlyeq_A y$ and the following hold:*

$$\breve{\Upsilon}_j(y) \leq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } y + v_j \in \mathcal{X} \setminus (K_A + x), \qquad (5.5)$$

*and*

$$\breve{\Upsilon}_j(y) \geq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } x + v_j \in \mathcal{X} \text{ and } y \notin K_A + x + v_j. \qquad (5.6)$$

*Then, for each $u \in [0,1]$,*

$$\Phi_\lambda(x, u) \preccurlyeq_A \breve{\Phi}_\lambda(y, u). \qquad (5.7)$$

*Proof.* First, we note that $\Phi_\lambda, \breve{\Phi}_\lambda$ have the following property: for every $u \in [0,1]$ and $1 \leq j \leq n$,

$$\text{if } \Phi_\lambda(x, u) = x + v_j, \quad \text{then } \breve{\Phi}_\lambda(y, u) \in \{y, y + v_j\}, \qquad (5.8)$$

since $I_j(x), \breve{I}_j(y) \subseteq [\frac{j-1}{n}, \frac{j}{n})$. Similarly,

$$\text{if } \breve{\Phi}_\lambda(y, u) = y + v_j, \quad \text{then } \Phi_\lambda(x, u) \in \{x, x + v_j\}. \qquad (5.9)$$

Furthermore, if $\breve{\Upsilon}_j(y) \geq \Upsilon_j(x)$, then

$$\Phi_\lambda(x, u) = x + v_j \text{ implies that } \breve{\Phi}_\lambda(y, u) = y + v_j, \qquad (5.10)$$

since under this condition, $I_j(x) \subseteq \breve{I}_j(y)$. Similarly, if $\breve{\Upsilon}_j(y) \leq \Upsilon_j(x)$, then

$$\breve{\Phi}_\lambda(y, u) = y + v_j \text{ implies that } \Phi_\lambda(x, u) = x + v_j. \qquad (5.11)$$

Now, to prove (5.7), fix $u \in [0,1]$. We consider two cases.

**Case 1:** $\breve{\Phi}_\lambda(y, u) = y + v_j$ for some $1 \leq j \leq n$.

Fix such an index $j$. Then, by (5.9), either $\Phi_\lambda(x, u) = x + v_j$ or $\Phi_\lambda(x, u) = x$.

a) If $\Phi_\lambda(x, u) = x + v_j$, then, by (3.1), $x + v_j \preccurlyeq_A y + v_j$ and therefore $\Phi_\lambda(x, u) \preccurlyeq_A \breve{\Phi}_\lambda(y, u)$.

b) If $\Phi_\lambda(x, u) = x$, then $y + v_j \in K_A + x$. To see this, we note that $y + v_j \in \mathcal{X}$ by (3.3) and since $\breve{\Upsilon}_j(y) > 0$ because $\breve{I}_j(y) \neq \emptyset$. Then, if $y + v_j \notin K_A + x$, by (5.5), we would have $\breve{\Upsilon}_j(y) \leq \Upsilon_j(x)$, which would imply that $\Phi_\lambda(x, u) = x + v_j$ by (5.11). But this contradicts the assumption that $\Phi_\lambda(x, u) = x$. Thus, $y + v_j \in K_A + x$ and so $\Phi_\lambda(x, u) = x \preccurlyeq_A y + v_j = \breve{\Phi}_\lambda(y, u)$.

**Case 2:** $\breve{\Phi}_\lambda(y, u) = y$. Again, we consider two subcases.

a) If $\Phi_\lambda(x, u) = x$, then (5.7) holds, since $x \preccurlyeq_A y$ by assumption.

26

b) If $\Phi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$, then $y \in K_A + x + v_j$ for the corresponding value of $j$. To see this, fix the value of $j$ for which $\Phi_\lambda(x, u) = x + v_j$ and notice that $x + v_j \in \mathcal{X}$ by (3.3) and since $\Upsilon_j(x) > 0$. If $y \notin K_A + x + v_j$, then by (5.6) we would have $\Upsilon_j(x) \leq \check{\Upsilon}_j(y)$, which would imply that $\check{\Phi}_\lambda(y, u) = y + v_j$. This contradicts the assumption that $\check{\Phi}_\lambda(y, u) = y$. Thus, we must have $y \in K_A + x + v_j$ and then $\Phi_\lambda(x, u) = x + v_j \preccurlyeq_A y = \check{\Phi}_\lambda(y, u)$.

$\square$

Now that all these preliminaries have been established under assumption (5.1), we proceed with the main part of the proof of Theorem 3.1 with this assumption. For this proof, we assume that all of the conditions of Theorem 3.1 hold and in addition that condition (5.1) holds. The latter ensures that the pair of continuous-time Markov chains with infinitesimal generators $Q$ and $\check{Q}$ are *uniformizable* (see Chapter 2 in Keilson [21]). With $\lambda > 0$ as in (5.2), the (possibly infinite) matrices [4] $P_\lambda(Q) := \frac{1}{\lambda}Q + I$ and $P_\lambda(\check{Q}) := \frac{1}{\lambda}\check{Q} + I$ are stochastic [5], where $I = (I_{x,y})_{x,y \in \mathcal{X}}$ is the identity matrix. Indeed, for $x \in \mathcal{X}$, $(P_\lambda(Q))_{x,x} = \frac{Q_{x,x}}{\lambda} + 1 = 1 - \frac{|Q_{x,x}|}{\lambda} \in [1 - \frac{1}{n}, 1]$, for $y \neq x$, $(P_\lambda(Q))_{x,y} = \frac{Q_{x,y}}{\lambda} \in [0, \frac{1}{n}]$ and $\sum_{y \in \mathcal{X}} (P_\lambda(Q))_{x,y} = \sum_{y \in \mathcal{X}} \frac{1}{\lambda}Q_{x,y} + 1 = 1$.

Now, let $x^\circ, \check{x}^\circ \in \mathcal{X}$ be such that $x^\circ \preccurlyeq_A \check{x}^\circ$. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where the following are defined:

(i) A Poisson process $N = \{N(t),\, 0 \leq t < \infty\}$ of rate $\lambda > 0$.

(ii) A sequence of independent and identically distributed (i.i.d.) random variables $U = (U_k)_{k \geq 1}$ where each $U_k$ has the uniform distribution on $[0, 1]$.

Additionally, choose $N$ to be independent of $U$. We construct two discrete-time processes, $Y = (Y_k)_{k \geq 0}$ and $\check{Y} = (\check{Y}_k)_{k \geq 0}$, by defining $Y_0 := x^\circ$, $\check{Y}_0 := \check{x}^\circ$, and for $k \geq 0$,

$$Y_{k+1} := \Phi_\lambda(Y_k, U_{k+1}), \qquad \check{Y}_{k+1} := \check{\Phi}_\lambda(\check{Y}_k, U_{k+1}). \tag{5.12}$$

Then $Y$ and $\check{Y}$ are discrete-time Markov chains with transition matrices $P_\lambda(Q)$ and $P_\lambda(\check{Q})$, respectively. Now, define the processes

$$X(t) := Y_{N(t)}, \qquad \check{X}(t) := \check{Y}_{N(t)}, \qquad t \geq 0. \tag{5.13}$$

According to Section 2.1 in Keilson [21] (see the discussion around Equation 2.1.6), $X$ and $\check{X}$ are continuous-time Markov chains with infinitesimal generators $Q$ and $\check{Q}$ respectively, and with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$.

In order to prove (3.7), it suffices to check that the following holds:

$$\mathbb{P}[Y_k \preccurlyeq_A \check{Y}_k] = 1, \qquad \text{for every } k \geq 0. \tag{5.14}$$

Indeed, if this is true, then $\mathbb{P}[Y_k \preccurlyeq_A \check{Y}_k \text{ for every } k \geq 0] = 1$ and therefore $\mathbb{P}[Y_{N(t)} \preccurlyeq_A \check{Y}_{N(t)}$ for every $t \geq 0] = 1$. We will prove (5.14) by induction on $k$. We already know that $x^\circ \preccurlyeq_A \check{x}^\circ$ and so (5.14) holds for $k = 0$. Now, assume $\mathbb{P}[Y_k \preccurlyeq_A \check{Y}_k] = 1$ for some $k \geq 0$.

---

[4]These "matrices" may have countably many rows and columns, in which case they could be considered as operators on $\ell^\infty$. For convenience, we still call them matrices here.

[5]Stochastic here means that all entries take values in $[0, 1]$ and all row sums equal one.

Since conditions (3.5) and (3.6) hold for every $x, y \in \mathcal{X}$ such that $x \preceq_A y$, by Lemma 5.1 we obtain that on a set of probability one, on which $Y_k \preceq_A \check{Y}_k$,

$$Y_{k+1} = \Phi_\lambda(Y_k, U_{k+1}) \preceq_A \check{\Phi}_\lambda(\check{Y}_k, U_{k+1}) = \check{Y}_{k+1}, \tag{5.15}$$

and so (5.14) holds with $k+1$ in place of $k$. This completes the induction step and so Theorem 3.1 is proved whenever (5.1) holds.

For the case where (5.1) does not hold, we construct the corresponding continuous-time Markov chains as a limit in distribution of appropriately coupled continuous-time Markov chains with truncated propensity functions for which (5.1) holds. Many elements for this case are similar to the previous case, although the use of Lemma 5.1 is different. We provide the details below, where we assume that the hypotheses of Theorem 3.1 hold.

We consider truncations of the propensity functions $\Upsilon$ and $\check{\Upsilon}$. More concretely, for $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preceq_A \check{x}^\circ$, let $M_0 \geq 1$ be an integer such that $\|x^\circ\|_\infty, \|\check{x}^\circ\|_\infty \leq M_0$. For every integer $M \geq M_0$, consider the finite set $\mathcal{X}_M := \{x \in \mathcal{X} \mid \|x\|_\infty \leq M\}$, together with the functions $\Upsilon_j^M, \check{\Upsilon}_j^M : \mathcal{X} \longrightarrow \mathbb{R}_+$ defined by $\Upsilon_j^M(x) := \Upsilon_j(x)\mathbb{1}_{\mathcal{X}_M}(x)$ and $\check{\Upsilon}_j^M(x) := \check{\Upsilon}_j(x)\mathbb{1}_{\mathcal{X}_M}(x)$ for $1 \leq j \leq n$ and $x \in \mathcal{X}$. We see that for every $M \geq M_0$, (3.3) holds with the functions $\Upsilon^M = (\Upsilon_1^M, \ldots, \Upsilon_n^M)$ and $\check{\Upsilon}^M = (\check{\Upsilon}_1^M, \ldots, \check{\Upsilon}_n^M)$ in place of $\Upsilon$ and $\check{\Upsilon}$. Also, since $\mathcal{X}_M$ is a finite set, $\sup_{x \in \mathcal{X}} \Upsilon_j^M(x) = \sup_{x \in \mathcal{X}_M} \Upsilon_j(x) < \infty$ and $\sup_{x \in \mathcal{X}} \check{\Upsilon}_j^M(x) = \sup_{x \in \mathcal{X}_M} \check{\Upsilon}_j(x) < \infty$ for every $1 \leq j \leq n$. Furthermore, by (3.5) and (3.6), we have that for every pair $x, y \in \mathcal{X}_M$ such that $x \preceq_A y$,

$$\check{\Upsilon}_j^M(y) \leq \Upsilon_j^M(x), \quad \text{for every } 1 \leq j \leq n \text{ such that } y + v_j \in \mathcal{X} \setminus (K_A + x), \text{ and}$$
$$\check{\Upsilon}_j^M(y) \geq \Upsilon_j^M(x), \quad \text{for every } 1 \leq j \leq n \text{ such that } x + v_j \in \mathcal{X} \text{ and } y \notin K_A + x + v_j. \tag{5.16}$$

Let $Q^M$ and $\check{Q}^M$ denote the infinitesimal generators associated with $\Upsilon^M$ and $\check{\Upsilon}^M$ respectively. We define an increasing sequence $\{\lambda_M\}_{M \geq M_0}$ of positive numbers such that $\lambda_M \longrightarrow \infty$ as $M \longrightarrow \infty$ and $\lambda_M > n \max \left\{ \sup_{x \in \mathcal{X}} \sum_{j=1}^n \Upsilon_j^M(x), \sup_{x \in \mathcal{X}} \sum_{j=1}^n \check{\Upsilon}_j^M(x) \right\}$ for every $M \geq M_0$. Define $\Phi_{\lambda_M}(\cdot, \cdot), \check{\Phi}_{\lambda_M}(\cdot, \cdot) : \mathcal{X} \times [0,1] \longrightarrow \mathcal{X}$ as in (5.4), but with $\Upsilon^M$ and $\check{\Upsilon}^M$ in place of $\Upsilon$ and $\check{\Upsilon}$, respectively. Since (5.16) holds, applying Lemma 5.1 with $\Upsilon^M, \check{\Upsilon}^M, \lambda_M, \Phi_{\lambda_M}, \check{\Phi}_{\lambda_M}$ in place of $\Upsilon, \check{\Upsilon}, \lambda, \Phi_\lambda, \check{\Phi}_\lambda$ yields that

$$\Phi_{\lambda_M}(x, u) \preceq_A \check{\Phi}_{\lambda_M}(y, u) \quad \text{for every } x, y \in \mathcal{X}_M \text{ such that } x \preceq_A y \text{ and } u \in [0,1]. \tag{5.17}$$

Now, for each $M \geq M_0$ consider a probability space $(\Omega^M, \mathcal{F}^M, \mathbb{P}^M)$ where the following are defined:

(i) A Poisson process $N^M = \{N^M(t), 0 \leq t < \infty\}$ of rate $\lambda_M > 0$.

(ii) An i.i.d. sequence $U^M = (U_k^M)_{k \geq 1}$ of uniform $[0,1]$ random variables.

Additionally, choose $N^M$ to be independent of $U^M$. For every $M \geq M_0$, we construct two discrete-time processes, $Y^M = (Y_k^M)_{k \geq 0}$ and $\check{Y}^M = (\check{Y}_k^M)_{k \geq 0}$, by defining $Y_0^M := x^\circ$, $\check{Y}_0^M := \check{x}^\circ$ and for $k \geq 0$,

$$Y_{k+1}^M := \Phi_{\lambda_M}(Y_k^M, U_{k+1}^M), \qquad \check{Y}_{k+1}^M := \check{\Phi}_{\lambda_M}(\check{Y}_k^M, U_{k+1}^M). \tag{5.18}$$

28

Similarly to the previous case, $Y^M$ and $\check{Y}^M$ are discrete-time Markov chains with transition matrices $P_{\lambda_M}(Q^M) := \frac{1}{\lambda_M}Q^M + I$ and $P_{\lambda_M}(\check{Q}^M) := \frac{1}{\lambda_M}\check{Q}^M + I$, respectively.

Now, we claim that for each $M \geq M_0$:

$$\mathbb{P}^M\left[Y^M_{k\wedge S^M} \preccurlyeq_A \check{Y}^M_{k\wedge S^M} \text{ for every } k \geq 0\right] = 1. \tag{5.19}$$

where $S^M := \inf\{k \geq 0 \mid Y^M_k \notin \mathcal{X}_M \text{ or } \check{Y}^M_k \notin \mathcal{X}_M\}$. In fact, (5.19) is equivalent to proving that $\mathbb{P}^M\left[Y^M_{k\wedge S^M} \preccurlyeq_A \check{Y}^M_{k\wedge S^M}\right] = 1$ for every $k \geq 0$, which we do by induction. We already know that $Y^M_0 \preccurlyeq_A \check{Y}^M_0$. Assuming the statement is true for some $k \geq 0$, to establish it for $k+1$ we distinguish between two cases. First, on $\{S^M \leq k\}$, $Y^M_{(k+1)\wedge S^M} = Y^M_{k\wedge S^M} \preccurlyeq_A \check{Y}^M_{k\wedge S^M} = \check{Y}^M_{(k+1)\wedge S^M}$, $\mathbb{P}^M$-a.s.. Second, on $\{S^M > k\}$, $Y^M_k \in \mathcal{X}_M, \check{Y}^M_k \in \mathcal{X}_M$, and by the induction assumption, $Y^M_k \preccurlyeq_A \check{Y}^M_k$, $\mathbb{P}^M$-a.s.. Applying Lemma 5.1, we obtain $\mathbb{P}^M$-a.s. on $\{S^M > k\}$ that

$$Y^M_{(k+1)\wedge S^M} = Y^M_{k+1} = \Phi_{\lambda_M}(Y^M_k, U^M_{k+1}) \preccurlyeq_A \check{\Phi}_{\lambda_M}(\check{Y}^M_k, U^M_{k+1}) = \check{Y}^M_{(k+1)\wedge S^M}, \tag{5.20}$$

where we have used (5.17).

Now, for each $M \geq M_0$, we define the processes

$$X^M(t) := Y^M_{N^M(t)}, \qquad \check{X}^M(t) := \check{Y}^M_{N^M(t)}, \qquad t \geq 0. \tag{5.21}$$

Then, $X^M$ and $\check{X}^M$ are continuous-time Markov chains with infinitesimal generators $Q^M$ and $\check{Q}^M$ respectively, and with initial conditions $X^M(0) = x^\circ$ and $\check{X}^M(0) = \check{x}^\circ$. Define $T^M := \inf\{t \geq 0 \mid X^M(t) \notin \mathcal{X}_M \text{ or } \check{X}^M(t) \notin \mathcal{X}_M\}$ and, because $Y^M$ and $\check{Y}^M$ are the discrete time skeletons for $X^M$ and $\check{X}^M$, we have that $\mathbb{P}^M$-a.s.

$$T^M = \inf\{t \geq 0 \mid N^M(t) = S^M\}. \tag{5.22}$$

Then, it follows from (5.19) that

$$\mathbb{P}^M\left[X^M(t \wedge T^M) \preccurlyeq_A \check{X}^M(t \wedge T^M) \text{ for every } t \geq 0\right] = 1. \tag{5.23}$$

We now prove that for every $t \geq 0$,

$$\mathbb{P}^M[T^M < t] \longrightarrow 0, \quad \text{as } M \longrightarrow \infty. \tag{5.24}$$

For this, let $T^M_{X^M} := \inf\{t \geq 0 \mid X^M(t) \notin \mathcal{X}_M\}$ and $T^M_{\check{X}^M} := \inf\{t \geq 0 \mid \check{X}^M(t) \notin \mathcal{X}_M\}$. Since $T^M = T^M_{X^M} \wedge T^M_{\check{X}^M}$, then

$$\mathbb{P}^M[T^M < t] \leq \mathbb{P}^M[T^M_{X^M} < t] + \mathbb{P}^M[T^M_{\check{X}^M} < t], \qquad \text{for every } t \geq 0. \tag{5.25}$$

Now, since $Q^M_{x,y} = Q_{x,y}$ for $x \in \mathcal{X}_M$ and $y \in \mathcal{X}$, $X^M(\cdot \wedge T^M_{X^M})$ will have the same distribution as a Markov chain with infinitesimal generator $Q$ and initial condition $x^\circ$, stopped at the first time it leaves $\mathcal{X}_M$. Because of this, $T^M_{X^M}$ has the same distribution as the first time a continuous-time Markov chain with infinitesimal generator $Q$ leaves $\mathcal{X}_M$. Since a continuous-time Markov chain with infinitesimal generator $Q$ has been assumed to not explode in finite time, we obtain that $\mathbb{P}^M[T^M_{X^M} < t] \longrightarrow 0$ as $M \to \infty$. Similar reasoning holds for $T^M_{\check{X}^M}$. Combining with (5.25), we obtain (5.24).

Denote by $\mathcal{D}([0,\infty),\mathcal{X}^2)$ the space of right-continuous functions from $[0,\infty)$ into $\mathcal{X}^2$ that also have finite left-limits. As usual, this space is endowed with Skorokhod's $J_1$ topology. The pair $(X^M, \check{X}^M)$ have paths in $\mathcal{D}([0,\infty),\mathcal{X}^2)$ and we obtain $(X, \check{X})$ as a limit in distribution of $(X^M, \check{X}^M)$ as $M \to \infty$. We first verify that the sequence of processes $\{(X^M, \check{X}^M)\}_{M \geq M_0}$ is tight. For this, it suffices to check that each sequence $\{X^M\}_{M \geq M_0}$ and $\{\check{X}^M\}_{M \geq M_0}$ is tight, which we do by means of Theorem 7.2 in Chapter 3 of Ethier & Kurtz [12]. Condition $(a)$ there (compact containment) is satisfied, because of (5.24) and because for $\tilde{M} \geq M \geq M_0$ we have that $X^{\tilde{M}}(\cdot \wedge T^M_{X^{\tilde{M}}})$ under $\mathbb{P}^{\tilde{M}}$ has the same law as $X^M(\cdot \wedge T^M_{X^M})$ under $\mathbb{P}^M$, where $T^M_{X^{\tilde{M}}} := \inf\{t \geq 0 \mid X^{\tilde{M}}(t) \notin \mathcal{X}_M\}$. To verify condition $(b)$ in Theorem 7.2 of [12], for $t_0 > 0$ fixed and $\eta > 0$, let $M_\eta \geq M_0$ be such that $\mathbb{P}^M[T^M_{X^M} < t_0] \leq \frac{\eta}{2}$ for all $M \geq M_\eta$. Then,

$$\mathbb{P}^M[w'(X^M, \delta, t_0) \geq \eta] \leq \mathbb{P}^M[w'(X^M, \delta, t_0) \geq \eta \,;\, T^M_{X^M} \geq t_0] + \mathbb{P}^M[T^M_{X^M} < t_0]$$
$$\leq \tilde{\mathbb{P}}[w'(\tilde{X}, \delta, t_0) \geq \eta \,;\, \tau^M_{\tilde{X}} \geq t_0] + \frac{\eta}{2}$$
$$\leq \tilde{\mathbb{P}}[w'(\tilde{X}, \delta, t_0) \geq \eta] + \frac{\eta}{2},$$

where $w'(\cdot, \cdot, \cdot)$ is the modulus of continuity, as defined in Equation (6.2), Chapter 3 of [12], $\tilde{X}$ under $\tilde{\mathbb{P}}$ is a realization of the Markov chain associated with the infinitesimal generator $Q$ that starts with $x^\circ$, and $\tau^M_{\tilde{X}} := \inf\{t \geq 0 \mid \tilde{X}(t) \notin \mathcal{X}_M\}$. Since $\tilde{X}$ under $\tilde{\mathbb{P}}$ is a single process with right-continuous paths having finite left-limits, the tightness applies to it and so the term $\tilde{\mathbb{P}}[w'(\tilde{X}, \delta, t_0) \geq \eta]$ can be made less than $\frac{\eta}{2}$ by choosing $\delta$ sufficiently small and so condition $(b)$ of Theorem 7.2 of [12] is satisfied. It follows that $\{X^M\}_{M \geq M_0}$ is tight. Similar reasoning yields tightness for $\{\check{X}^M\}_{M \geq M_0}$.

It follows that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two processes $X$ and $\check{X}$ defined there, having paths that are right-continuous with finite left-limits, and a subsequence $\{M_k\}_{k \geq 1}$ such that $M_k \to \infty$ as $k \to \infty$, and the sequence $\{(X^{M_k}, \check{X}^{M_k})\}_{k \geq 1}$ converges in distribution to the pair of processes $(X, \check{X})$. To identify the law of the limit, note that since $\{Q^{M_k}\}_{k \geq 1}$ converges pointwise to $Q$, for any function $f$ with bounded support in $\mathcal{X}$, $f(X(t)) - \int_0^t Qf(X(s))ds$ will inherit the martingale property of $f(X^{M_k}(t)) - \int_0^t Q^{M_k} f(X^{M_k}(s))ds$. It follows from the martingale characterization that $X$ is a continuous-time Markov chain with infinitesimal generator $Q$ (see Chapter 4 in Ethier & Kurtz [12]). Similarly, $\check{X}$ will be a continuous-time Markov chain with infinitesimal generator $\check{Q}$. In addition, the processes have inherited initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$.

Finally, to show that (3.7) holds, consider the set

$$F = \{(f, g) \in \mathcal{D}([0,\infty),\mathcal{X}^2) \mid f(t) \preccurlyeq_A g(t) \text{ for all } t \geq 0\}, \tag{5.26}$$

which is closed in the Skorokhod topology. From (5.23) we know that the stopped processes satisfy $\mathbb{P}^{M_k}[(X^{M_k}(\cdot \wedge T^{M_k}), \check{X}^{M_k}(\cdot \wedge T^{M_k})) \in F] = 1$ for every $k \geq 1$. Furthermore, from (5.24) we know that $T^{M_k} \longrightarrow \infty$ in probability as $k \to \infty$. The reader may verify that this last fact, along with the convergence of $(X^{M_k}, \check{X}^{M_k})$ to $(X, \check{X})$, implies that $(X^{M_k}(\cdot \wedge T^{M_k}), \check{X}^{M_k}(\cdot \wedge T^{M_k}))$ converges in distribution to $(X, \check{X})$ as $k \to \infty$. By the Portmanteau Theorem (see Theorem 2.1 in Billingsley [6]),

$$1 = \limsup_{k \to \infty} \mathbb{P}^{M_k}[(X^{M_k}(\cdot \wedge T^{M_k}), \check{X}^{M_k}(\cdot \wedge T^{M_k})) \in F] \leq \mathbb{P}[(X, \check{X}) \in F] \tag{5.27}$$

and we obtain (3.7).

**Remark 5.1.** The proof of Theorem 3.1 provides a method to simulate the sample paths for the continuous-time Markov chains $X$ and $\breve{X}$ in a coupled manner for the case where (5.1) holds. Roughly speaking, the procedure consists of determining $\lambda > 0$ as in (5.2), $\Phi_\lambda, \breve{\Phi}_\lambda$ as in (5.4), $Y, \breve{Y}$ as in (5.12) and $X, \breve{X}$ as in (5.13). For the benefit of the reader, this method is described as an algorithm in Section S.4, which yields coupled sample paths under the assumptions of Theorem 3.2, 3.3 and S.2.

## 5.2  Proof of Theorem 3.2

By Theorem 3.1, it suffices to prove that for every $x, y \in \mathcal{X}$ such that $x \preccurlyeq_A y$, conditions (3.5) and (3.6) hold. For this, we make some observations first. Consider $x, y \in \mathcal{X}$ such that $x \preccurlyeq_A y$ and let $1 \leq j \leq n$. Observe that $x \preccurlyeq_A y + v_j$ will hold if and only if $A(y + v_j - x) \geq 0$ which is equivalent to:

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m. \tag{5.28}$$

Similarly, $x + v_j \preccurlyeq_A y$ will hold if and only if

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m. \tag{5.29}$$

Since $x \preccurlyeq_A y$, then $\langle A_{i\bullet}, y - x \rangle \geq 0$ for every $1 \leq i \leq m$. Now, consider $i \in \{1, ..., m\}$ such that $\langle A_{i\bullet}, y - x \rangle > 0$. Since $A \in \mathbb{Z}^{m \times d}$ and $y - x \in \mathbb{Z}^d$, then $\langle A_{i\bullet}, y - x \rangle \geq 1$. This yields that

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 1 + \langle A_{i\bullet}, v_j \rangle \geq 0, \tag{5.30}$$

since $\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. Similarly, $\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 1 - \langle A_{i\bullet}, v_j \rangle \geq 0$. By observing that the interior of $K_A + x$ is of the form $\text{int}(K_A + x) = \{y \in \mathbb{R}^d \mid Ax < Ay\}$, the latter argument shows that for every $x \in \mathcal{X}$ and $y \in \text{int}(K_A + x) \cap \mathcal{X}$, we have

$$x \preccurlyeq_A y + v_j \text{ and } x + v_j \preccurlyeq_A y, \qquad \text{for every } 1 \leq j \leq n. \tag{5.31}$$

Now, lets check condition (3.5). For this, let $x, y \in \mathcal{X}$ be such that $x \preccurlyeq_A y$ and let $1 \leq j \leq n$ be such that $y + v_j \in \mathcal{X} \backslash (K_A + x)$. By (5.31), $y \notin \text{int}(K_A + x)$ and since $y \in K_A + x$, we must have $y \in \partial(K_A + x) = \{z \in K_A + x \mid \langle A_{i\bullet}, z \rangle = \langle A_{i\bullet}, x \rangle \text{ for some } 1 \leq i \leq m\}$, the boundary of $K_A + x$. Consider the set of indices $\mathbf{K}_y := \{i \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle, 1 \leq i \leq m\}$, which is non-empty. Observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and from (5.30), $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - x \rangle = \langle A_{i\bullet}, v_j \rangle$. From this, we can infer that there exists an $i_k \in \mathbf{K}_y$ such that $\langle A_{i_k\bullet}, v_j \rangle < 0$. Indeed, if this was not the case, then $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$ for every $i \in \mathbf{K}_y$ and consequently (5.28) would hold. This contradicts the fact that $y + v_j \notin K_A + x$. By (3.9), we know that $\langle A_{i_k\bullet}, v_j \rangle < 0$ implies $\breve{\Upsilon}_j(y) \leq \Upsilon_j(x)$ and we conclude that (3.5) holds.

To check condition (3.6), let $x, y \in \mathcal{X}$ be such that $x \preccurlyeq_A y$ and let $1 \leq j \leq n$ be such that $x + v_j \in \mathcal{X}$ and $y \notin K_A + x + v_j$. Again, by (5.31), we obtain that $y \in \partial(K_A + x)$ and $\mathbf{K}_y \neq \emptyset$. For every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle = -\langle A_{i\bullet}, v_j \rangle$. From this, we can infer that there exists an $i_k \in \mathbf{K}_y$ such that $\langle A_{i_k\bullet}, v_j \rangle > 0$. By (3.10), we know that $\langle A_{i_k\bullet}, v_j \rangle > 0$ implies $\breve{\Upsilon}_j(y) \geq \Upsilon_j(x)$ and we conclude that (3.6) holds.

31

## 5.3 Proof of Theorem 3.3

The proof of this result uses similar general ideas to the ones used in the proof of Theorem 3.1. However, since the conditions involve sums, the construction is somewhat different and more complex and we provide the details below. Let us consider again a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\breve{\Upsilon} = (\breve{\Upsilon}_1, \ldots, \breve{\Upsilon}_n)$ such that (3.3) holds. In the following, let $A \in \mathbb{Z}^{m \times d}$ be a matrix with non-zero rows such that condition (i) of Theorem 3.3 holds.

We initially assume that $\sup_{x \in \mathcal{X}} \Upsilon_j(x) < \infty$ and $\sup_{x \in \mathcal{X}} \breve{\Upsilon}_j(x) < \infty$ for every $1 \leq j \leq n$, and let $\lambda > 0$ such that (5.2) holds. We shall relax these assumptions later. We start by defining functions analogous to $\Phi_\lambda$ and $\breve{\Phi}_\lambda$ as defined in (5.4), although this time, the construction is more involved.

Recall that $s$ denotes the size of the set $\{Av_j \mid 1 \leq j \leq n\}$ and that the index sets $G^k \neq \emptyset$, $1 \leq k \leq s$, defined in (3.12), are such that $Av_j = \eta^k$ for all $j \in G^k$, $1 \leq k \leq s$. Consider a bijection $\sigma : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ such that the vectors $v_{\sigma(1)}, \ldots, v_{\sigma(n)}$ have the property that the first $|G^1|$ vectors have indices in $G^1$, the next $|G^2|$ vectors have indices in $G^2$, and so on. More precisely, the bijection $\sigma$ is such that for $1 \leq k \leq s$, $Av_{\sigma(q)} = \eta^k$, whenever $\sum_{\ell=1}^{k-1} |G^\ell| + 1 \leq q \leq \sum_{\ell=1}^{k} |G^\ell|$. Recall for this that a sum over an empty set is taken to equal zero.

For $x \in \mathcal{X}$, we define a family of intervals $\{I^k(x) \mid 1 \leq k \leq s\}$ as follows. Let $p_0 := 0$, and for $1 \leq k \leq s$, inductively define $p_k := \sum_{\ell=1}^{k} |G^\ell|$, and

$$I^k(x) := \bigcup_{q=p_{k-1}+1}^{p_k} I_q^k(x), \tag{5.32}$$

where for $p_{k-1} + 1 \leq q \leq p_k$,

$$I_q^k(x) := \left[ \frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^{q-1} \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda}, \frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^{q} \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda} \right). \tag{5.33}$$

The sets $I_q^k(x)$, with $1 \leq k \leq s$ and $p_{k-1} + 1 \leq q \leq p_k$, are mutually disjoint, and by (5.2), the length of $I^k(x)$ is less than $\frac{p_k - p_{k-1}}{n} = \frac{|G^k|}{n}$, and so the sum of the lengths of $\{I^k(x) \mid 1 \leq k \leq s\}$ is less than $\frac{1}{n} \sum_{k=1}^{s} |G^k| = 1$. Now, let us define $\Psi_\lambda(\cdot, \cdot) : \mathcal{X} \times [0,1] \longrightarrow \mathcal{X}$ by

$$\Psi_\lambda(x, u) := x + \sum_{k=1}^{s} \sum_{q=p_{k-1}+1}^{p_k} v_{\sigma(q)} \mathbb{1}_{I_q^k(x)}(u), \qquad x \in \mathcal{X}, \ u \in [0,1]. \tag{5.34}$$

Note that $Av_{\sigma(q)} = \eta^k$ for $p_{k-1} + 1 \leq q \leq p_k$, $1 \leq k \leq s$. From the above properties of $I_q^k(x)$, we have that for any $u \in [0,1]$, either $u \notin \bigcup_{k=1}^{s} I^k(x)$ or $u \in I_q^k(x)$ for exactly one $k$ and $q$ such that $I_q^k(x) \neq \emptyset$. The latter condition implies, by (5.33), that $\Upsilon_{\sigma(q)}(x) > 0$ and then, by (3.3), $x + v_{\sigma(q)} \in \mathcal{X}$. This shows that $\Psi_\lambda(\cdot, \cdot)$ is well-defined as an $\mathcal{X}$-valued function.

In an analogous manner to that above, we can define intervals $\breve{I}^k(x)$, $\breve{I}_q^k(x)$, $1 \leq k \leq s$, $p_{k-1} + 1 \leq q \leq p_k$, $x \in \mathcal{X}$ and a function $\breve{\Psi}_\lambda : \mathcal{X} \times [0,1] \longrightarrow \mathcal{X}$, as in (5.32) – (5.34), but with $\breve{\Upsilon}_j(x)$, $\breve{I}^k(x)$, $\breve{I}_q^k(x)$, $\breve{\Psi}_\lambda$ in place of $\Upsilon_j(x)$, $I^k(x)$, $I_q^k(x)$, $\Psi_\lambda$.

32

**Lemma 5.2.** *Suppose that $x, y \in \mathcal{X}$ are such that $x \preccurlyeq_A y$ and the following hold: whenever $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$, we have*

$$\sum_{j \in G^k} \breve{\Upsilon}_j(y) \leq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for every } k \text{ such that } \eta_i^k < 0, \tag{5.35}$$

*and*

$$\sum_{j \in G^k} \breve{\Upsilon}_j(y) \geq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for every } k \text{ such that } \eta_i^k > 0. \tag{5.36}$$

*Then, for each $u \in [0, 1]$,*

$$\Psi_\lambda(x, u) \preccurlyeq_A \breve{\Psi}_\lambda(y, u). \tag{5.37}$$

*Proof.* First, we note that $\Psi_\lambda, \breve{\Psi}_\lambda$ have the following properties: for every $u \in [0, 1]$, $1 \leq k \leq s$, $j \in G^k$,

$$\text{if } \Psi_\lambda(x, u) = x + v_j, \text{ then } \breve{\Psi}_\lambda(y, u) \in \{y + v_\ell : \ell \in G^k\} \cup \{y\}, \tag{5.38}$$

since $I^k_{\sigma^{-1}(j)}(x), \breve{I}^k_{\sigma^{-1}(\ell)}(y) \subseteq [\frac{p_{k-1}}{n}, \frac{p_k}{n})$ for $\ell \in G^k$. Similarly,

$$\text{if } \breve{\Psi}_\lambda(y, u) = y + v_j, \text{ then } \Psi_\lambda(x, u) \in \{x + v_\ell : \ell \in G^k\} \cup \{x\}. \tag{5.39}$$

Furthermore, for $1 \leq k \leq s$, $j \in G^k$, if $\sum_{\ell \in G^k} \breve{\Upsilon}_\ell(y) \geq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, then

$$\Psi_\lambda(x, u) = x + v_j \text{ implies that } \breve{\Psi}_\lambda(y, u) = y + v_\ell \text{ for some } \ell \in G^k, \tag{5.40}$$

since under the condition, $I^k(x) \subseteq \breve{I}^k(y)$. Similarly, if $\sum_{\ell \in G^k} \breve{\Upsilon}_\ell(y) \leq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, then

$$\breve{\Psi}_\lambda(y, u) = y + v_j \text{ implies that } \Psi_\lambda(x, u) = x + v_\ell \text{ for some } \ell \in G^k. \tag{5.41}$$

We also have that, for $1 \leq k \leq s$ and $j \in G^k$, $x \preccurlyeq_A y + v_j$ if and only if

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \tag{5.42}$$

Similarly, $x + v_j \preccurlyeq_A y$ if and only if

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \tag{5.43}$$

Furthermore, for $1 \leq k \leq s$ and $j, \ell \in G^k$, since $Av_j = Av_\ell$ and $x \preccurlyeq_A y$, then

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, (v_j - v_\ell) \rangle = \langle A_{i\bullet}, y - x \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \tag{5.44}$$

To prove (5.37), we first consider the situation where $y \in \text{int}(K_A + x) = \{w \in \mathbb{R}^d \mid Ax < Aw\}$. Then, for each $1 \leq i \leq m$, $\langle A_{i\bullet}, y - x \rangle > 0$ and since $A \in \mathbb{Z}^{m \times d}$ and $y - x \in \mathbb{Z}^d$, we have $\langle A_{i\bullet}, y - x \rangle \geq 1$. This implies that for $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 1 + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m, \tag{5.45}$$

since $\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$ by condition $(i)$ of Theorem 3.3. Similarly, for $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 1 - \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \tag{5.46}$$

It follows from (5.44) – (5.46) that if $y \in \text{int}(K_A + x) \cap \mathcal{X}$, then for any $1 \leq k \leq s$ and $j, \ell \in G^k$:

$$x \preccurlyeq_A y + v_j, \ x + v_j \preccurlyeq_A y \ \text{ and } \ x + v_\ell \preccurlyeq_A y + v_j. \tag{5.47}$$

We also have, by assumption, that $x \preccurlyeq_A y$. It follows that if $y \in \text{int}(K_A + x) \cap \mathcal{X}$, then $\{x, x + v_\ell \mid \ell \in G^k\} \preccurlyeq_A \{y, y + v_j \mid j \in G^k\}$ for $1 \leq k \leq s$ and consequently (5.37) holds for all $u \in [0, 1]$.

Now, we turn to the other situation where $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$. Then $\mathbf{K}_y := \{i \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle, 1 \leq i \leq m\}$ is non-empty. Let $u \in [0, 1]$. We consider two cases.

**Case 1:** $\breve{\Psi}_\lambda(y, u) = y + v_j$ for some $1 \leq j \leq n$.

Fix such an index $j$. Consider the unique $1 \leq k \leq s$ such that $j \in G^k$. Then, by (5.39), either $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, or $\Psi_\lambda(x, u) = x$.

a) If $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, then, since $x \preccurlyeq_A y$ and $Av_j = Av_\ell$, we have $x + v_\ell \preccurlyeq_A y + v_j$. Hence, $\Psi_\lambda(x, u) \preccurlyeq_A \breve{\Psi}_\lambda(y, u)$ and (5.37) holds.

b) If $\Psi_\lambda(x, u) = x$, we claim that $y + v_j \in K_A + x$. To see this, observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (5.45), $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - x \rangle = \langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = -1$, then by (5.35), we would have $\sum_{\ell \in G^k} \breve{\Upsilon}_\ell(y) \leq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, which would imply that $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, but this contradicts the assumption that $\Psi_\lambda(x, u) = x$. So we must have $\langle A_{i\bullet}, v_j \rangle \geq 0$ and hence $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, $y + v_j \in K_A + x$ and so $\Psi_\lambda(x, u) = x \preccurlyeq_A y + v_j = \breve{\Psi}_\lambda(y, u)$ holds.

**Case 2:** $\breve{\Psi}_\lambda(y, u) = y$. Again, we consider two subcases.

a) If $\Psi_\lambda(x, u) = x$, then (5.37) holds, because $x \preccurlyeq_A y$.

b) If $\Psi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$, we claim that $y \in K_A + x + v_j$ for the corresponding value of $j$. To see this, fix the value of $j$ for which $\Psi_\lambda(x, u) = x + v_j$, let $1 \leq k \leq s$ be such that $j \in G^k$, and observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (5.46), $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle = -\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = 1$, then by (5.36), we would have $\sum_{\ell \in G^k} \breve{\Upsilon}_\ell(y) \geq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, which would imply that $\breve{\Psi}_\lambda(y, u) = y + v_\ell$ for some $\ell \in G^k$. This would contradict the assumption that $\breve{\Psi}_\lambda(y, u) = y$. So we must have $\langle A_{i\bullet}, v_j \rangle \leq 0$ and hence $\langle A_{i\bullet}, y - (x + v_j) \rangle = \langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, we have $y \in K_A + x + v_j$ and then $\Psi_\lambda(x, u) = x + v_j \preccurlyeq_A y = \breve{\Psi}_\lambda(y, u)$. $\qquad \square$

In order to prove Theorem 3.3, from here on we can follow a similar procedure to the one used in the proof of Theorem 3.1 after Lemma 5.1 was proved there. For the case where (5.1) holds, we define two discrete-time processes, $Y = (Y_k)_{k \geq 0}$ and $\breve{Y} = (\breve{Y}_k)_{k \geq 0}$, by defining $Y_0 := x^\circ$, $\breve{Y}_0 := \breve{x}^\circ$, and for $k \geq 0$,

$$Y_{k+1} := \Psi_\lambda(Y_k, U_{k+1}), \qquad \breve{Y}_{k+1} := \breve{\Psi}_\lambda(\breve{Y}_k, U_{k+1}), \tag{5.48}$$

and define $X$ and $\breve{X}$ using these and an independent Poisson process $N$ as in (5.13). For the case where (5.1) does not hold, we can use a truncation procedure similar to that for Theorem 3.1. In both cases, we use Lemma 5.2 instead of Lemma 5.1.

# 6 Conclusion

In this work, we first reviewed the concept of Stochastic Chemical Reaction Networks (SCRNs), a class of continuous-time Markov chain models frequently used to describe the stochastic behavior of chemical reaction systems. We also gave the definitions of preorder and increasing set considered in this paper. In Section 3.2, we presented the main theoretical results of this paper. We first derived, by exploiting uniformization and then coupling of stochastic processes (see Grassmann [14] and Keilson [21]), three theorems which give practical sufficient conditions for stochastic dominance of one continuous-time Markov chain over another. More precisely, these theorems provide conditions under which, when one or more parameters is changed monotonically, the system is almost surely "higher" with respect to a certain preorder. While the first theorem (Theorem 3.1) can be used for any SCRN, it has extensive conditions to check. The second set of theorems (Theorems 3.2, 3.3) can be used for more specific SCRN classes, but they have assumptions that only need to be checked at the boundary of certain translated convex cones. All these theorems can be applied to SCRNs with either finite or countably many states. In Section 3.3, we exploited these tools to develop two theorems to specifically study the monotonicity properties of stationary distributions and mean first passage times depending on system parameters.

Subsequently, in Section 4, we presented some illustrative examples to highlight the advantages of using our theoretical tools in order to study the stochastic behavior of SCRNs. Specifically, we focused on two common models for enzymatic kinetics (see Michaelis & Menten [23], Kang et al. [20], Del Vecchio & Murray [10] and Anderson et al. [2]), on a model inspired by Braess's paradox (see Calvert et al. [9]) and on a recently developed model describing the main interactions among histone modifications alone, and together with an expressed protein (see Bruno et al. [8]). In these illustrative examples we see that our sufficient conditions can be easy to check and our results can be also used to study networks with a countably infinite number of states. Furthermore, the conclusions obtained by using our theorems are true for trajectories of the Markov chains, yielding results for both transient and steady state behavior.

Overall, in this paper we derived and presented theorems that can be used for the theoretical study of monotonicity of SCRNs associated to a variety of chemical reaction systems. Future work will include the adaptation of our theoretical tools to other forms of monotonicity for SCRNs (see Definition 5.1.1 in Muller & Stoyan [24] as an example), the investigation of possible correlations between the network graph properties and the monotonicity properties of the SCRN (extension of the work of Angeli et al. [5] to SCRNs), and the application of our results to deterministic chemical reaction network through appropriate limits.

**Supplementary information (SI) file:** file containing detailed mathematical derivations for some of our examples, a generalization of Theorem 3.3, and an algorithm for coupled stochastic simulation.

**Ethics declarations - conflict of interest:** The authors declare that they have no con-

flicts of interest.

**Data availability:** Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

# References

[1] ALLIS, C. D. AND CAPARROS, M.-L. AND JENUWEIN, T. AND REINBERG, D. (2015). *Epigenetics, 2nd edition.* Cold Spring Harbor Laboratory Press.

[2] ANDERSON, D. F., CRACIUN, G., AND KURTZ, T. G. (2010). Product-form stationary distributions for deficiency zero chemical reaction networks. *Bulletin of Mathematical Biology*, 72:1947-–1970.

[3] ANDERSON, D.F. AND KURTZ, T.G. (2015). *Stochastic Analysis of Biochemical Systems.* Springer International Publishing.

[4] ANGELI, D., AND SONTAG, E. D. (2003). Monotone control systems. *IEEE Transactions on Automatic Control*, 48(10): 1684–1698.

[5] ANGELI, D., DE LEEHEER, P., AND SONTAG, E. D. (2006). On the structural monotonicity of chemical reaction networks. *Proceedings of the 45th IEEE Conference on Decision & Control*, 7–12.

[6] BILLINGSLEY, P. (1999). *Convergence of Probability Measures, 2nd edition.* Wiley Series in Probability and Statistics.

[7] BRAESS, D. (1968). Über ein paradoxon aus der verkehrsplanung.*Unternehmensforschung Operations Research - Recherche Opérationnelle*, 27(1):258-–268.

[8] BRUNO, S., WILLIAMS, R.J., AND DEL VECCHIO, D. (2022). Epigenetic cell memory: The gene's inner chromatin modification circuit.*PLOS Computational Biology*, Public Library of Science, 18(4):1–27.

[9] CALVERT, B. SOLOMON, W., AND ZIEDINS, I. (1997). Braess's paradox in a queueing network with state-dependent routing. *Journal of Applied Probability*, 34(1):134—154.

[10] DEL VECCHIO, D., AND MURRAY, R.M. (2014). *Biomolecular Feedback Systems.* Princeton University Press.

[11] DODD, I. B., MICHEELSEN, M. A., SNEPPEN, K., AND THON, G. (2007). Theoretical analysis of epigenetic cell memory by nucleosome modification.*Cell*, 129(4):813–822.

[12] ETHIER, S. AND KURTZ, T. (1986). *Markov Processes: Characterization and Convergence.* John Wiley, New York.

[13] GORI, G., MILAZZO, P., AND NASTI, L. (2019). Towards an efficient verification method for monotonicity properties of chemical reaction networks. *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies*, 250—257.

[14] GRASSMANN, W.K. (1977). Transient solutions in Markovian queueing systems. *Computers & Operations Research*, 4(1):47-53.

[15] GUNAWAN, R., CAO, Y., PETZOLD, L., AND DOYLE, F.J. (2005). Sensitivity analysis of discrete stochastic systems.*Biophysical Journal*, 88(4):2530-–2540.

[16] GUPTA, A., AND KHAMMASH, M. (2014). Sensitivity analysis for stochastic chemical reaction networks with multiple time-scales. *Electronic Journal of Probability*, 19: 1–53.

[17] HASELTINE, E.L., AND RAWLINGS J.B. (2002). Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics.*The Journal of Chemical Physics*, 117:6959–6969.

[18] HIRSCH, M., AND SMITH, H. (2006). Chapter 4 Monotone Dynamical Systems. In A. Canada, P. Drabek, and A. Fonda (Eds.), *Handbook of Differential Equations: Ordinary Differential Equations; Vol.2.* North-Holland, 239–357.

[19] KAMAE, T., KRENGEL, U., AND O'BRIEN, G.L. (1977). Stochastic inequalities on partially ordered spaces. *The Annals of Probability.* 5(6): 899–912.

[20] KANG, H. W., KHUDABUKHSH, W.R., KOEPPL, H., AND REMPALA, G.A. (2019). Quasi-steady-state approximations derived from the stochastic model of enzyme kinetics. *Bulletin of Mathematical Biology*, 81: 1303—1336.

[21] KEILSON, J. (1979). *Markov Chain Models — Rarity and Exponentiality.* Applied Mathematical Sciences, Springer, New York, NY.

[22] MASSEY, W.A. (1987). Stochastic orderings for Markov processes on partially ordered spaces. *Mathematics of Operations Research*, 12(2):350—367.

[23] MICHAELIS, L., AND MENTEN, M.L. (1913). Die kinetik der invertinwirkung. *Biochem. Z.* 49:333—369.

[24] MULLER, A., AND STOYAN, D. (2002). *Comparison Methods for Stochastic Models and Risks.* Wiley Series in Probability and Statistics.

[25] SRIVASTAVA, R., YOU, L., SUMMERS, J. AND YIN, J. (2002). Stochastic vs. deterministic modeling of intracellular viral kinetics. *Journal of Theoretical Biology,* 218(3):309–321.

# Comparison Theorems for Stochastic Chemical Reaction Networks

Felipe A. Campos[1,*], Simone Bruno[2,*], Yi Fu[1], Domitilla Del Vecchio[2], and Ruth J. Williams[1]

[1] *Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0112. Email:* (`fcamposv,yif064,rjwilliams`)`@ucsd.edu`

[2] *Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. Emails:* (`sbruno,ddv`)`@mit.edu`

[*] *These authors contributed equally: F. A. Campos and S. Bruno*

## Supplementary Information (SI) file

### S.1 Criteria for Positive Recurrence and Exponential Ergodicity with Application to Examples 4.2 and 4.5

#### S.1.1 Foster-Lyapunov Conditions for Positive Recurrence and Exponential Ergodicity

Here we recall fairly general conditions for positive recurrence and exponential ergodicity of a continuous-time Markov chain. Such conditions are well known and are usually referred to as Foster-Lyapunov-type conditions. We also apply these to the Markov chains in Examples 4.2 and 4.5, in subsections S.1.2 and S.1.3.

**Theorem S.1.** *Let $X$ be an irreducible continuous-time Markov chain with state space $\mathcal{X}$ and infinitesimal generator $Q$. Suppose $V : \mathcal{X} \to \mathbb{R}_+$ is norm-like, that is $\{x \in \mathcal{X} : V(x) \leq a\}$ is compact [6] for each $a \in \mathbb{R}_+$. Further assume that for some $c > 0$, $d > 0$ and a compact set $C$,*

$$QV(x) \leq -c + d\mathbb{1}_C(x), \quad \text{for all } x \in \mathcal{X}. \tag{S.1}$$

*Then, $X$ is non-explosive and positive recurrent, and has a unique stationary distribution $\pi$. If instead of (S.1), we have that for some $c' > 0$ and $d' > 0$,*

$$QV(x) \leq -c'V(x) + d', \quad \text{for all } x \in \mathcal{X}, \tag{S.2}$$

*then (S.1) automatically holds and the consequences stated above hold, and in addition, the stationary distribution satisfies*

$$\pi(V) = \sum_{x \in \mathcal{X}} \pi_x V(x) < \infty,$$

*and there is $0 < B < \infty$ and $0 < \beta < 1$ such that for all $t \geq 0$ and $x \in \mathcal{X}$,*

$$\sum_{y \in \mathcal{X}} |P_{xy}(t) - \pi_y| \leq \|P_{x\bullet}(t) - \pi\|_{V+1} \leq B(V(x) + 1)\beta^t, \tag{S.3}$$

---

[6] Since $X$ is a Markov chain, $\mathcal{X}$ is finite or countable, and we endow it with the usual discrete topology consisting of all subsets of $\mathcal{X}$.

*where $P_{xy}(t) = \mathbb{P}[X(t) = y | X(0) = x]$ and*

$$\|P_{x\bullet}(t) - \pi\|_{V+1} = \sup_{|g| \le V+1} \left| \sum_{y \in \mathcal{X}} (P_{xy}(t) - \pi_y)g(y) \right|.$$

**Remark S.1.** When the second inequality in (S.3) holds, we say that the Markov chain is exponentially ergodic in the $(V + 1)$-norm.

*Proof.* We will verify the sufficient conditions for each of non-explosion, positive recurrence and exponential ergodicity given in Meyn & Tweedie [1]. Note that $X$ is a Borel right-process under the definition in Sharpe [2]. In addition, since the state space is discrete, each compact set is finite and therefore petite.

For $m \in \mathbb{Z}_+$, if $Q_m$ is the infinitesimal generator for the Markov chain $X$ killed upon exit from $O_m = \{x \in \mathcal{X} : V(x) \le m\}$ [7], then $Q_m V(x) \le QV(x)$ for $x \in O_m$. It then follows from (S.1) that conditions (CD0) and (CD2) (with $f = 1$) in [1] hold with $Q_m$ in place of $\mathscr{A}_m$ there. By Theorem 2.1 and Theorem 4.2 in [1], the Markov chain is non-explosive and positive recurrent, and it has a unique stationary distribution $\pi$.

On the other hand, if (S.2) holds, then (S.1) holds, using the norm like property of $V$. Furthermore, (S.2) implies that conditions (CD0), (CD2) (with $f = V + 1$) and (CD3) in [1] hold with $Q_m$ in place of $\mathscr{A}_m$ there. By Theorem 2.1, Theorem 4.2 and Theorem 6.1 in [1], the Markov chain is non-explosive and positive recurrent, with a unique stationary distribution $\pi$ such that $\pi(V) < \infty$, and it is exponentially ergodic in the $(V + 1)$-norm, that is the second inequality in (S.3) holds for all $t \ge 0$ and $x \in \mathcal{X}$. For fixed $t \ge 0$ and $x \in \mathcal{X}$, setting $g(y) = \text{sgn}(P_{xy}(t) - \pi_y)$, for $y \in \mathcal{X}$, we have that $|g| \le 1 \le V + 1$, and

$$\sum_{y \in \mathcal{X}} |P_{xy}(t) - \pi_y| = \left| \sum_{y \in \mathcal{X}} (P_{xy}(t) - \pi_y)g(y) \right| \le \|P_{x\bullet}(t) - \pi\|_{V+1},$$

yielding the first inequality in (S.3). $\qquad\square$

### S.1.2 Application to Example 4.2

For Example 4.2, we first show that the Markov chain is irreducible. For this, consider $x^\circ = (0, 0, \text{E}_{\text{tot}}, 0)$ and any fixed state $x = (x_1, x_2, x_3, \text{E}_{\text{tot}} - x_3) : 0 \le x_3 \le \text{E}_{\text{tot}}$. Starting at $x^\circ$, by having reaction ⑤ fire $x_1 + \text{E}_{\text{tot}} - x_3 + x_2$ times in succession, then having reaction ①, immediately followed by reaction ③, fire $x_2$ times in succession and then reaction ① fire $\text{E}_{\text{tot}} - x_3$ times, without any other reactions firing, we see that the Markov chain can transition with positive probability from $x^\circ$ to $x$. Since each reaction is reversible, it also follows that the Markov chain can transition from $x$ to $x^\circ$ with positive probability. Thus, the Markov chain is irreducible.

Next we will introduce a norm-like function $V$ and show that (S.1) holds. For each $x \in \mathcal{X}$, let

$$V(x) = x_1^2 + ((2\text{E}_{\text{tot}} - 1)b + 1)x_2 + bx_4^2,$$

---

[7]Upon exit from $O_m$, the killed process goes to a cemetery state $\Delta_m$ in $\mathcal{X} \backslash O_m$ where $V(\Delta_m) = \min\{V(x) : x \in \mathcal{X} \backslash O_m\}$.

where

$$b = \frac{1 + (\kappa_5 + \kappa_2 \mathrm{E}_{\mathrm{tot}} + \kappa_3 \mathrm{E}_{\mathrm{tot}}) + \frac{(2\kappa_5 + \kappa_6 + 2\kappa_2 \mathrm{E}_{\mathrm{tot}})^2}{8\kappa_6}}{\kappa_2 \mathrm{E}_{\mathrm{tot}}(2\mathrm{E}_{\mathrm{tot}} - 1)}.$$

(S.4)

Notice that $b > 0$ since $\mathrm{E}_{\mathrm{tot}} \geq 1$. Then, for each $a \in \mathbb{R}_+$, $\{x \in \mathcal{X} : V(x) \leq a\}$ consists of finitely many states, and for any $x \in \mathcal{X}$,

$$
\begin{aligned}
QV(x) &= \sum_{j=1}^{6} \Upsilon_j(x) \cdot (V(x + v_j) - V(x)) \\
&= \kappa_1 x_1 x_3 \cdot (((x_1 - 1)^2 + b(x_4 + 1)^2) - (x_1^2 + b x_4^2)) \\
&\quad + \kappa_2 x_4 \cdot (((x_1 + 1)^2 + b(x_4 - 1)^2) - (x_1^2 + b x_4^2)) \\
&\quad + \kappa_3 x_4 \cdot ((((2\mathrm{E}_{\mathrm{tot}} - 1)b + 1)(x_2 + 1) + b(x_4 - 1)^2) - (((2\mathrm{E}_{\mathrm{tot}} - 1)b + 1)x_2 + b x_4^2)) \\
&\quad + \kappa_4 x_2 x_3 \cdot ((((2\mathrm{E}_{\mathrm{tot}} - 1)b + 1)(x_2 - 1) + b(x_4 + 1)^2) - (((2\mathrm{E}_{\mathrm{tot}} - 1)b + 1)x_2 + b x_4^2)) \\
&\quad + \kappa_5 \cdot ((x_1 + 1)^2 - x_1^2) + \kappa_6 x_1 \cdot ((x_1 - 1)^2 - x_1^2) \\
&= \kappa_1 x_1 x_3 \cdot (-2x_1 + 1 + b(2x_4 + 1)) + \kappa_2 x_4 \cdot (2x_1 + 1 + b(-2x_4 + 1)) \\
&\quad + \kappa_3 x_4 \cdot (((2\mathrm{E}_{\mathrm{tot}} - 1)b + 1) + b(-2x_4 + 1)) \\
&\quad + \kappa_4 x_2 x_3 \cdot (-((2\mathrm{E}_{\mathrm{tot}} - 1)b + 1) + b(2x_4 + 1)) \\
&\quad + \kappa_5 \cdot (2x_1 + 1) + \kappa_6 x_1 \cdot (-2x_1 + 1) \\
&= -(2\kappa_1 x_3 + 2\kappa_6) \cdot x_1^2 + (2\kappa_5 + \kappa_6 + \kappa_1(1 + b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4) \cdot x_1 \\
&\quad + \kappa_4(-2b\mathrm{E}_{\mathrm{tot}} x_3 + 2b x_3 x_4 + (2b - 1)x_3) \cdot x_2 \\
&\quad + \kappa_5 + (\kappa_2(1 + b) + \kappa_3(2b\mathrm{E}_{\mathrm{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2 \\
&= -(2\kappa_1 x_3 + 2\kappa_6) \cdot x_1^2 + (2\kappa_5 + \kappa_6 + \kappa_1(1 + b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4) \cdot x_1 \\
&\quad + \kappa_4(-2b x_3^2 + (2b - 1)x_3) \cdot x_2 + \kappa_5 + (\kappa_2(1 + b) + \kappa_3(2b\mathrm{E}_{\mathrm{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2.
\end{aligned}
$$

The last equality uses the fact that $x_3 = \mathrm{E}_{\mathrm{tot}} - x_4$. For the following, we note that $-2b x_3^2 + (2b - 1)x_3 = (-2b x_3 + (2b - 1))x_3 \leq -1$ when $x_3 \geq 1$.

We now consider two cases for $QV(x)$: when $x_3 = 0$ and $x_3 > 0$. For the first case, when $x_3 = 0$, we have $x_4 = \mathrm{E}_{\mathrm{tot}}$ and

$$QV(x) = -2\kappa_6 \cdot x_1^2 + (2\kappa_5 + \kappa_6 + 2\kappa_2 \mathrm{E}_{\mathrm{tot}}) \cdot x_1 + \kappa_5 + (\kappa_2(1 + b) + \kappa_3)\mathrm{E}_{\mathrm{tot}} - 2b\kappa_2 \mathrm{E}_{\mathrm{tot}}^2.$$

As a quadratic function, the last expression is bounded above by

$$b\kappa_2 \mathrm{E}_{\mathrm{tot}}(1 - 2\mathrm{E}_{\mathrm{tot}}) + (\kappa_5 + \kappa_2 \mathrm{E}_{\mathrm{tot}} + \kappa_3 \mathrm{E}_{\mathrm{tot}}) + \frac{(2\kappa_5 + \kappa_6 + 2\kappa_2 \mathrm{E}_{\mathrm{tot}})^2}{8\kappa_6} = -1,$$

since $b$ is chosen as in (S.4). For the second case when $x_3 \in \{1, \ldots, \mathrm{E}_{\mathrm{tot}}\}$, we have

$$
\begin{aligned}
QV(x) &\leq -2\kappa_6 \cdot x_1^2 + (2\kappa_5 + \kappa_6 + \kappa_1(1 + b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4) \cdot x_1 \\
&\quad - \kappa_4 \cdot x_2 + \kappa_5 + (\kappa_2(1 + b) + \kappa_3(2b\mathrm{E}_{\mathrm{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2 \\
&\leq -2\kappa_6 \cdot x_1^2 + \kappa_6 \cdot x_1^2 + \left( \frac{2\kappa_5 + \kappa_6 + \kappa_1(1 + b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4}{2\sqrt{\kappa_6}} \right)^2 \\
&\quad - \kappa_4 \cdot x_2 + \kappa_5 + (\kappa_2(1 + b) + \kappa_3(2b\mathrm{E}_{\mathrm{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2 \\
&\leq -\kappa_6 \cdot x_1^2 - \kappa_4 \cdot x_2 + \frac{(2\kappa_5 + \kappa_6 + \kappa_1(1 + b)\mathrm{E}_{\mathrm{tot}} + 2b\kappa_1 \mathrm{E}_{\mathrm{tot}}^2 + 2\kappa_2 \mathrm{E}_{\mathrm{tot}})^2}{4\kappa_6} \\
&\quad + \kappa_5 + (\kappa_2(1 + b) + \kappa_3(2b\mathrm{E}_{\mathrm{tot}} + 1))\mathrm{E}_{\mathrm{tot}},
\end{aligned}
$$

where we have used the fact that $2ab \leq a^2 + b^2$ for the second inequality and the last expression will be less than or equal to $-1$ whenever $x_1$ or $x_2$ is sufficiently large.

3

Let $C = \{x \in \mathcal{X} : QV(x) > -1\}$. Then, by the above, $C$ consists of finitely many points, which implies that $C$ is a compact set. Then, (S.1) holds with $c = 1$ and $d = 1 + \max_{x \in C} QV(x) \vee 0$. It follows by Theorem S.1 that the Markov chain in Example 4.2 is non-explosive and positive recurrent, and has a unique stationary distribution $\pi$.

### S.1.3 Application to Example 4.5

For Example 4.5, we first check that the Markov chain is irreducible. For this, consider $x^\circ = (0, D_{\text{tot}}, 0)$ and any fixed state $x = (x_1, x_2, x_3) : 0 \le x_1 + x_2 \le D_{\text{tot}}$. Starting at $x^\circ$, by having reaction ⑤ fire $x_3$ times, then reaction ③ fire $D_{\text{tot}} - x_2$ times, and finally, reaction ② fire $x_1$ times, without any other reactions firing, we see that the Markov chain can transition from $x^\circ$ to $x$ with positive probability. For the reverse transition, by having reaction ⑥ fire $x_3$ times, then reaction ④ fire $x_1$ times, and finally, reaction ① fire $D_{\text{tot}} - x_2$ times, we see that the Markov chain can transition from $x$ to $x^\circ$ with positive probability. Thus, the Markov chain is irreducible.

Next we will introduce a norm-like function $V$ and show that (S.2) holds. For each $x \in \mathcal{X}$, let
$$V(x) = x_3.$$
Then $\{x \in \mathcal{X} : V(x) \le a\} = \{x \in \mathbb{Z}_+^3 : x_1 + x_2 \le D_{\text{tot}}, x_3 \le a\}$ consists of finitely many states for each $a \in \mathbb{R}_+$, and for any $x \in \mathcal{X}$,

$$
\begin{aligned}
QV(x) &= \sum_{j=1}^{6} \Upsilon_j(x) \cdot (V(x + v_j) - V(x)) = \sum_{j=1}^{6} \Upsilon_j(x) \cdot V(v_j) \\
&= \Upsilon_1(x) \cdot 0 + \Upsilon_2(x) \cdot 0 + \Upsilon_3(x) \cdot 0 + \Upsilon_4(x) \cdot 0 + \kappa_{5a} x_2 \cdot 1 + \kappa_{6a} x_3 \cdot (-1) \\
&\le -\kappa_{6a} x_3 + \kappa_{5a} D_{\text{tot}} = -c' V(x) + d',
\end{aligned}
$$

where $c' = \kappa_{6a}$ and $d' = \kappa_{5a}$. Therefore, we conclude by Theorem S.1 that the Markov chain in Example 4.5 is non-explosive and positive recurrent with a unique stationary distribution $\pi$ such that $\pi(V) < \infty$, and it is exponentially ergodic in the $(V + 1)$-norm.

## S.2 Derivation of Markov chain transition directions, $v_j$

### S.2.1 Example 4.2

The set of reactions associated to the chemical reaction system in Fig. 3(a) is given by $\mathcal{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+), (v_5^-, v_5^+), (v_6^-, v_6^+)\}$, where $(v_1^-, v_1^+)$, $(v_2^-, v_2^+)$, $(v_3^-, v_3^+)$ are defined as in Example 4.1, and $v_4^- = (0, 1, 1, 0)^T$, $v_4^+ = (0, 0, 0, 1)^T$, $v_5^- = v_6^+ = (0, 0, 0, 0)^T$, $v_5^+ = v_6^- = (1, 0, 0, 0)^T$. Then, the potential transitions of the Markov chain are in six possible directions, $v_j = v_j^+ - v_j^-$ for $j = 1, ..., 6$, where $v_1 = -v_2 = (-1, 0, -1, 1)^T$, $v_3 = -v_4 = (0, 1, 1, -1)^T$, and $v_5 = -v_6 = (1, 0, 0, 0)^T$.

### S.2.2 Example 4.3

The set of five reactions associated to the chemical reaction system in Fig. 4(a) is given by $\mathcal{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+), (v_5^-, v_5^+)\}$, where $v_1^- = v_3^- = (1, 0, 0, 0)^T$, $v_1^+ = v_2^- = v_5^- = (0, 1, 0, 0)^T$, $v_3^+ = v_4^- = v_5^+ = (0, 0, 1, 0)^T$, $v_2^+ = v_4^+ = (0, 0, 0, 1)^T$. Then, the potential transitions of the Markov chain are in five possible directions, $v_j = v_j^+ - v_j^-$, $j =$

1, ..., 5, where $v_1 = (-1, 1, 0, 0)^T$, $v_2 = (0, -1, 0, 1)^T$, $v_3 = (-1, 0, 1, 0)^T$, $v_4 = (0, 0, -1, 1)^T$ and $v_5 = (0, -1, 1, 0)^T$.

## S.3  A generalization of Theorem 3.3

Here, we provide a more general version of Theorem 3.3. The simpler form given as Theorem 3.3 in the main text was used there because it is more straightforward to state and the conditions are easier to verify. However, the more general version of the theorem provided in this section can be useful in some cases, such as Example 4.3 (see Section S.3.2).

The generalization relies on the idea that grouping of vectors can be more general than what is described in (3.12), and so we introduce the following assumption.

**Assumption S.1.** *Consider a collection of distinct vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows. Suppose that there exists a partition* [8] *$\{G^1, \ldots, G^s\}$ of $\{1, \ldots, n\}$ and an associated bijection $\sigma : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ such that for $p_0 := 0$ and $1 \leq k \leq s$, with $p_k := \sum_{\ell=1}^{k} |G^\ell|$, we have $\sigma(q) \in G^k$ for $p_{k-1} + 1 \leq q \leq p_k$, and whenever $p_{k-1} + 2 \leq q \leq p_k$, we also have for each $1 \leq i \leq m$ that either $\langle A_{i\bullet}, v_{\sigma(q)} \rangle$ is equal to $\langle A_{i\bullet}, v_{\sigma(q-1)} \rangle$ or it is 0.*

**Theorem S.2.** *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, suppose that Assumption S.1 holds and consider two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n)$, such that (3.3) holds and the associated continuous-time Markov chains do not explode in finite time. Further suppose that both of the following conditions hold:*

*(i) For each $1 \leq j \leq n$, the vector $Av_j$ has entries in $\{-1, 0, 1\}$ only.*

*(ii) For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that for each $1 \leq k \leq s$,*

$$\sum_{j \in G_i^{k,-}} \check{\Upsilon}_j(y) \leq \sum_{j \in G_i^{k,-}} \Upsilon_j(x), \quad \text{where } G_i^{k,-} = \{j \in G^k \mid \langle A_{i\bullet}, v_j \rangle = -1\},$$

*and*

$$\sum_{j \in G_i^{k,+}} \check{\Upsilon}_j(y) \geq \sum_{j \in G_i^{k,+}} \Upsilon_j(x), \quad \text{where } G_i^{k,+} = \{j \in G^k \mid \langle A_{i\bullet}, v_j \rangle = 1\}.$$

*Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous-time Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators $Q$ and $\check{Q}$, associated with $\Upsilon$ and $\check{\Upsilon}$ respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$ and such that:*

$$\mathbb{P}\left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0\right] = 1. \tag{S.5}$$

**Remark S.2.** Under Assumption S.1, for a given $1 \leq k \leq s$ and $1 \leq i \leq m$, at most one of $G_i^{k,-}$ and $G_i^{k,+}$ is non-empty. If $G_i^{k,-} \neq \emptyset$, then $G_i^{k,-} = \{\sigma(q) \mid p_{k-1} + 1 \leq q \leq q^*\}$ where $q^* = \max\{q \mid p_{k-1} + 1 \leq q \leq p_k \text{ and } \langle A_{i\bullet}, v_{\sigma(q)} \rangle = -1\}$. On the other hand, if $G_i^{k,+} \neq \emptyset$, then $G_i^{k,+} = \{\sigma(q) \mid p_{k-1} + 1 \leq q \leq q^*\}$ where $q^* = \max\{q \mid p_{k-1} + 1 \leq q \leq p_k \text{ and } \langle A_{i\bullet}, v_{\sigma(q)} \rangle = 1\}$. Furthermore, in either case, for $q^* < q \leq p_k$, we have $\langle A_{i\bullet}, v_{\sigma(q)} \rangle = 0$.

---

[8] In particular, $\{G^1, \ldots, G^s\}$ is a finite collection of non-empty disjoint sets of distinct numbers, the union of which is $\{1, \ldots, n\}$.

The proof for Theorem S.2 can be found in Section S.3.1. With Theorem S.2 in place, we can extend Theorems 3.4 and 3.5 by adding an additional alternative condition $(iv)$: Assumption S.1 holds, and conditions $(i)$ and $(ii)$ in Theorem S.2 are satisfied. If this is satisfied instead of one of $(i)$ – $(iii)$ in Theorems 3.4 and 3.5, then the conclusions of these theorems about (mean) first passage times and stationary distributions will still hold.

### S.3.1 Proof of Theorem S.2

We initially assume that $\sup_{x\in\mathcal{X}}\Upsilon_j(x) < \infty$ and $\sup_{x\in\mathcal{X}}\breve{\Upsilon}_j(x) < \infty$ for every $1 \leq j \leq n$, and let $\lambda > 0$ such that (5.2) holds. We shall relax these assumptions later. Further suppose that Assumption S.1 and condition (i) of Theorem S.2 both hold. For $x \in \mathcal{X}$, define $I^k(x)$, $I_q^k(x)$, $\Psi_\lambda$, $\breve{I}^k(x)$, $\breve{I}_q^k(x)$ and $\breve{\Psi}_\lambda$ in the same manner as in the proof of Theorem 3.3 (see (5.32) – (5.34)), with $\{G^k \mid 1 \leq k \leq s\}$, $\{p_k \mid 0 \leq k \leq s\}$ and $\sigma$ as in Assumption S.1. Our proof of Theorem S.2 has some elements that are the same as those for the proof of Theorem 3.3. However, some additional elements are needed. We give the details for completeness. As for Theorem 3.3, $\Psi_\lambda(\cdot,\cdot)$ and $\breve{\Psi}_\lambda(\cdot,\cdot)$ are well-defined as $\mathcal{X}$-valued functions.

**Lemma S.1.** *Suppose that $x, y \in \mathcal{X}$ are such that $x \preccurlyeq_A y$ and the following hold: whenever $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$, we have that for each $1 \leq k \leq s$,*

$$\sum_{j\in G_i^{k,-}} \breve{\Upsilon}_j(y) \leq \sum_{j\in G_i^{k,-}} \Upsilon_j(x), \quad \text{where } G_i^{k,-} = \{j \in G^k \mid \langle A_{i\bullet}, v_j\rangle = -1\}, \quad (S.6)$$

*and*

$$\sum_{j\in G_i^{k,+}} \breve{\Upsilon}_j(y) \geq \sum_{j\in G_i^{k,+}} \Upsilon_j(x), \quad \text{where } G_i^{k,+} = \{j \in G^k \mid \langle A_{i\bullet}, v_j\rangle = 1\}. \quad (S.7)$$

*Then, for each $u \in [0,1]$:*

$$\Psi_\lambda(x,u) \preccurlyeq_A \breve{\Psi}_\lambda(y,u). \quad (S.8)$$

*Proof.* First, we note that $\Psi_\lambda, \breve{\Psi}_\lambda$ have the following properties: for every $u \in [0,1]$, $1 \leq k \leq s$, $j \in G^k$,

$$\text{if } \Psi_\lambda(x,u) = x + v_j, \text{ then } \breve{\Psi}_\lambda(y,u) \in \{y + v_\ell : \ell \in G^k\} \cup \{y\}, \quad (S.9)$$

since $I_{\sigma^{-1}(j)}^k(x), \breve{I}_{\sigma^{-1}(\ell)}^k(y) \subseteq [\frac{p_{k-1}}{n}, \frac{p_k}{n})$ for $\ell \in G^k$. Similarly,

$$\text{if } \breve{\Psi}_\lambda(y,u) = y + v_j, \text{ then } \Psi_\lambda(x,u) \in \{x + v_\ell : \ell \in G^k\} \cup \{x\}. \quad (S.10)$$

Furthermore, for $1 \leq k \leq s$, $1 \leq i \leq m$, $j \in G_i^{k,+}$, if $\sum_{\ell\in G_i^{k,+}} \breve{\Upsilon}_\ell(y) \geq \sum_{\ell\in G_i^{k,+}} \Upsilon_\ell(x)$, then

$$\Psi_\lambda(x,u) = x + v_j \text{ implies that } \breve{\Psi}_\lambda(y,u) = y + v_\ell \text{ for some } \ell \in G_i^{k,+}, \quad (S.11)$$

since under the condition, we have $\cup_{q=p_{k-1}+1}^{q^*} I_q^k(x) \subseteq \cup_{q=p_{k-1}+1}^{q^*} \breve{I}_q^k(y)$ where $q^* = \max\{q\mid p_{k-1}+1 \leq q \leq p_k \text{ and } \langle A_{i\bullet}, v_{\sigma(q)}\rangle = 1\}$ and, by Assumption S.1, $G_i^{k,+} = \{\sigma(q)\mid p_{k-1}+1 \leq q \leq q^*\}$. Similarly, for $1 \leq k \leq s$, $1 \leq i \leq m$, $j \in G_i^{k,-}$, if $\sum_{\ell\in G_i^{k,-}} \breve{\Upsilon}_\ell(y) \leq \sum_{\ell\in G_i^{k,-}} \Upsilon_\ell(x)$, then

$$\breve{\Psi}_\lambda(y,u) = y + v_j \text{ implies that } \Psi_\lambda(x,u) = x + v_\ell \text{ for some } \ell \in G_i^{k,-}. \quad (S.12)$$

6

We also have that, for $1 \leq k \leq s$ and $j \in G^k$, $x \preccurlyeq_A y + v_j$ if and only if

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m. \qquad \text{(S.13)}$$

Similarly, $x + v_j \preccurlyeq_A y$ if and only if

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m. \qquad \text{(S.14)}$$

Furthermore, for $1 \leq k \leq s$ and $j, \ell \in G^k$, $x + v_\ell \preccurlyeq_A y + v_j$ if and only if

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j - v_\ell \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m. \qquad \text{(S.15)}$$

To prove (S.8), we first consider the situation where $y \in \text{int}(K_A + x) = \{w \in \mathbb{R}^d \mid Ax < Aw\}$. Then, for each $1 \leq i \leq m$, $\langle A_{i\bullet}, y - x \rangle > 0$ and since $A \in \mathbb{Z}^{m \times d}$ and $y - x \in \mathbb{Z}^d$, we have $\langle A_{i\bullet}, y - x \rangle \geq 1$. This implies that for each $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 1 + \langle A_{i\bullet}, v_j \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m, \qquad \text{(S.16)}$$

since $\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$ by condition $(i)$ of Theorem S.2. Similarly, for each $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 1 - \langle A_{i\bullet}, v_j \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m. \qquad \text{(S.17)}$$

In addition, for $1 \leq k \leq s$ and $j, \ell \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j - v_\ell \rangle \geq 1 + \langle A_{i\bullet}, v_j - v_\ell \rangle \geq 0, \qquad \text{for every } 1 \leq i \leq m, \qquad \text{(S.18)}$$

since, by Assumption S.1, if $\langle A_{i\bullet}, v_j \rangle \neq 0$, then either $\langle A_{i\bullet}, v_\ell \rangle = \langle A_{i\bullet}, v_j \rangle$ or $\langle A_{i\bullet}, v_\ell \rangle = 0$. It follows from (S.16) – (S.18) that if $y \in \text{int}(K_A + x) \cap \mathcal{X}$, then for any $1 \leq k \leq s$ and $j, \ell \in G^k$:

$$x \preccurlyeq_A y + v_j, \ x + v_j \preccurlyeq_A y \ \text{and} \ x + v_\ell \preccurlyeq_A y + v_j. \qquad \text{(S.19)}$$

We also have, by assumption, that $x \preccurlyeq_A y$. It follows that if $y \in \text{int}(K_A + x) \cap \mathcal{X}$, then $\{x, x + v_\ell \mid \ell \in G^k\} \preccurlyeq_A \{y, y + v_j \mid j \in G^k\}$ for $1 \leq k \leq s$ and consequently (S.8) holds for all $u \in [0, 1]$.

Now, we turn to the other situation where $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$. Then $\mathbf{K}_y := \{i \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle, 1 \leq i \leq m\}$ is non-empty. Let $u \in [0, 1]$. We consider two cases.

**Case 1:** $\breve{\Psi}_\lambda(y, u) = y + v_j$ for some $1 \leq j \leq n$.

Fix such an index $j$. Consider the unique $1 \leq k \leq s$ such that $j \in G^k$. Then, by (S.10), either $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, or $\Psi_\lambda(x, u) = x$.

a) Suppose $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$. Observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (S.18), $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle = \langle A_{i\bullet}, v_j \rangle - \langle A_{i\bullet}, v_\ell \rangle$. For each $i \in \mathbf{K}_y$,

   i) if $\langle A_{i\bullet}, v_j \rangle = -1$ and $\langle A_{i\bullet}, v_\ell \rangle = 0$, then $j \in G_i^{k,-}$ and $\ell \notin G_i^{k,-}$. By (S.6), we would then have $\sum_{r \in G_i^{k,-}} \breve{\Upsilon}_r(y) \leq \sum_{r \in G_i^{k,-}} \Upsilon_r(x)$, which would imply by (S.12) that $\Psi_\lambda(x, u) = x + v_r$ for some $r \in G_i^{k,-}$. Since we are assuming that $\Psi_\lambda(x, u) = x + v_\ell$ and we know the vectors $v_1, \ldots, v_n$ are distinct, we obtain that $\ell = r \in G_i^{k,-}$. This contradicts $\ell \notin G_i^{k,-}$.

7

ii) if $\langle A_{i\bullet}, v_j \rangle = 0$ and $\langle A_{i\bullet}, v_\ell \rangle = 1$, then $j \notin G_i^{k,+}$ and $\ell \in G_i^{k,+}$. By (S.7), we would then have $\sum_{r \in G_i^{k,+}} \check{\Upsilon}_r(y) \geq \sum_{r \in G_i^{k,+}} \Upsilon_r(x)$, which would imply by (S.11) that $j \in G_i^{k,+}$. This contradicts $j \notin G_i^{k,+}$.

iii) in all the other cases, that is when

$$(\langle A_{i\bullet}, v_j \rangle, \langle A_{i\bullet}, v_\ell \rangle) \in \{(1,0), (0,-1), (1,1), (0,0), (-1,-1)\},$$

we have $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle \geq 0$.

Combining the above for case a), we see that $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle \geq 0$ for each $1 \leq i \leq m$, which implies that $\Psi_\lambda(x, u) = x + v_\ell \preccurlyeq_A y + v_j = \check{\Psi}_\lambda(y, u)$.

b) Suppose $\Psi_\lambda(x, u) = x$. We claim that $y + v_j \in K_A + x$. To see this, observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (S.16), $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - x \rangle = \langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = -1$, which means $j \in G_i^{k,-}$, then by (S.6) we would have $\sum_{\ell \in G_i^{k,-}} \check{\Upsilon}_\ell(y) \leq \sum_{\ell \in G_i^{k,-}} \Upsilon_\ell(x)$, which would imply by (S.12) that $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G_i^{k,-}$, but this contradicts the assumption that $\Psi_\lambda(x, u) = x$. So we must have $\langle A_{i\bullet}, v_j \rangle \geq 0$ and hence $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, $y + v_j \in K_A + x$ and so $\Psi_\lambda(x, u) = x \preccurlyeq_A y + v_j = \check{\Psi}_\lambda(y, u)$.

**Case 2:** $\check{\Psi}_\lambda(y, u) = y$. Again, we consider two subcases.

a) If $\Psi_\lambda(x, u) = x$, then (S.8) holds, because $x \preccurlyeq_A y$.

b) If $\Psi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$, we claim that $y \in K_A + x + v_j$ for the corresponding value of $j$. To see this, fix the value of $j$ for which $\Psi_\lambda(x, u) = x + v_j$, let $1 \leq k \leq s$ be such that $j \in G^k$, and observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (S.17), $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle = -\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = 1$, which means $j \in G_i^{k,+}$, then by (S.7), we would have $\sum_{\ell \in G_i^{k,+}} \check{\Upsilon}_\ell(y) \geq \sum_{\ell \in G_i^{k,+}} \Upsilon_\ell(x)$, which would imply by (S.11) that $\check{\Psi}_\lambda(y, u) = y + v_\ell$ for some $\ell \in G_i^{k,+}$, but this contradicts the assumption that $\check{\Psi}_\lambda(y, u) = y$. So we must have $\langle A_{i\bullet}, v_j \rangle \leq 0$ and hence $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, we have $y \in K_A + x + v_j$ and then $\Psi_\lambda(x, u) = x + v_j \preccurlyeq_A y = \check{\Psi}_\lambda(y, u)$. $\square$

In order to prove Theorem S.2, from here on we can follow a similar procedure to the one used in the proof of Theorem 3.1 after Lemma 5.1 was proved there. For the case where (5.1) holds, we define two discrete-time processes, $Y = (Y_k)_{k \geq 0}$ and $\check{Y} = (\check{Y}_k)_{k \geq 0}$, by defining $Y_0 := x^\circ$, $\check{Y}_0 := \check{x}^\circ$, and for $k \geq 0$,

$$Y_{k+1} := \Psi_\lambda(Y_k, U_{k+1}), \qquad \check{Y}_{k+1} := \check{\Psi}_\lambda(\check{Y}_k, U_{k+1}), \tag{S.20}$$

and define $X$ and $\check{X}$ using these and an independent Poisson process $N$ as in (5.13). For the case where (5.1) does not hold, we can use a truncation procedure similar to that for Theorem 3.1. In both cases, we use Lemma S.1 instead of Lemma 5.1.

### S.3.2 Two other $A$ matrices for Example 4.3

Let

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}. \tag{S.21}$$

For $x \in \mathcal{X}$, consider infinitesimal transition rates $\breve{\Upsilon}_1(x), \breve{\Upsilon}_2(x), \breve{\Upsilon}_3(x), \breve{\Upsilon}_4(x)$ and $\breve{\Upsilon}_5(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x)$ and $\Upsilon_5(x)$ in (4.7), but with $\breve{\kappa}_i$ in place of $\kappa_i$ where $\breve{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4$, and $\breve{\kappa}_5 \leq \kappa_5$. Suppose that $\kappa_2 > \kappa_4$. Now, let us verify that the assumptions of Theorem S.2 hold. Condition $(i)$ holds since $Av_1 = (1, 0, -1)^T$, $Av_2 = (0, 0, 1)^T$, $Av_3 = (1, -1, -1)^T$, $Av_4 = (0, 1, 1)^T$ and $Av_5 = (0, -1, 0)^T$. Assumption S.1 holds with $G^1 = \{3, 1\}$, $G^2 = \{4, 2\}$, $G^3 = \{5\}$ and $\sigma(1) = 3$, $\sigma(2) = 1$, $\sigma(3) = 4$, $\sigma(4) = 2$, $\sigma(5) = 5$. To verify that condition $(ii)$ of Theorem S.2 holds, fix $x \in \mathcal{X}$ and first consider $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_3 \geq w_3, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \leq \breve{\Upsilon}_1(y) + \breve{\Upsilon}_3(y)$. Since $y \in \partial_1(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 = \kappa_1 y_1 = \breve{\kappa}_1 y_1 = \breve{\Upsilon}_1(y)$ and $\Upsilon_3(x) = \kappa_3 x_1 = \kappa_3 y_1 = \breve{\kappa}_3 y_1 = \breve{\Upsilon}_3(y)$, and so the desired inequality holds with equality. Secondly, consider $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_3 = w_3, x_2 \geq w_2, x_4 \leq w_4\}$. Given that $\langle A_{2\bullet}, v_3 \rangle = \langle A_{2\bullet}, v_5 \rangle = -1$ and $\langle A_{2\bullet}, v_4 \rangle = 1$, we need to check that $\Upsilon_3(x) \geq \breve{\Upsilon}_3(y)$, $\Upsilon_4(x) \leq \breve{\Upsilon}_4(y)$ and $\Upsilon_5(x) \geq \breve{\Upsilon}_5(y)$. Since $y \in \partial_2(K_A + x) \cap \mathcal{X}$, then $\Upsilon_3(x) = \kappa_3 x_1 \geq \kappa_3 y_1 = \breve{\kappa}_3 y_1 = \breve{\Upsilon}_3(y)$, $\Upsilon_4(x) = \kappa_4 x_3 = \kappa_4 y_3 = \breve{\kappa}_4 y_3 = \breve{\Upsilon}_4(y)$ and $\Upsilon_5(x) = \kappa_5 x_2 \geq \kappa_5 y_2 \geq \breve{\kappa}_5 y_2 = \breve{\Upsilon}_5(y)$, and so the desired inequality holds. Lastly, consider $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_3 \geq w_3, x_2 + x_3 = w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = -1$ and $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_4 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \geq \breve{\Upsilon}_1(y) + \breve{\Upsilon}_3(y)$ and $\Upsilon_2(x) + \Upsilon_4(x) \leq \breve{\Upsilon}_2(y) + \breve{\Upsilon}_4(y)$. For $y \in \partial_3(K_A + x) \cap \mathcal{X}$, since $\kappa_2 > \kappa_4$ was assumed, we have that $\Upsilon_2(x) + \Upsilon_4(x) = \kappa_2 x_2 + \kappa_4 x_3 = (\kappa_2 - \kappa_4)x_2 + \kappa_4(x_2 + x_3) \leq (\kappa_2 - \kappa_4)y_2 + \kappa_4(y_2 + y_3) = \breve{\kappa}_2 y_2 + \breve{\kappa}_4 y_3 = \breve{\Upsilon}_2(y) + \breve{\Upsilon}_4(y)$ and $\Upsilon_1(x) = \kappa_1 x_1 \geq \kappa_1 y_1 = \breve{\kappa}_1 y_1 = \breve{\Upsilon}_1(y)$, $\Upsilon_3(x) = \kappa_3 x_1 \geq \kappa_3 y_1 = \breve{\kappa}_3 y_1 = \breve{\Upsilon}_3(y)$. Thus, the conditions of Theorem S.2 are satisfied and so the conclusion of that theorem holds.

Let $\Gamma = \{(0, 0, 0, S_{tot})\}$. This is an increasing set in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$. Let $T_{(0,0,0,S_{tot})}$, respectively $\breve{T}_{(0,0,0,S_{tot})}$ be the first time that the Markov chain $X$, respectively $\breve{X}$, reaches the set $\Gamma$. Then, by the generalization of Theorem 3.4, if $X(0) = \breve{X}(0) = (S_{tot}, 0, 0, 0)$, we have that $\breve{T}_{(0,0,0,S_{tot})} \preccurlyeq_{st} T_{(0,0,0,S_{tot})}$. It follows that increasing $\kappa_5$ will increase the mean first passage time from $(S_{tot}, 0, 0, 0)$ to $(0, 0, 0, S_{tot})$ when $\kappa_2 > \kappa_4$ (See Figure S.1). Indeed, when $\kappa_2 > \kappa_4$, it takes a longer time to get to $(0, 0, 0, S_{tot})$ from $(S_{tot}, 0, 0, 0)$ if reaction $\textcircled{5}$ is added to the system without that reaction.

On the other hand, suppose

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

and infinitesimal transition rates $\breve{\Upsilon}_1(x), \breve{\Upsilon}_2(x), \breve{\Upsilon}_3(x), \breve{\Upsilon}_4(x)$ and $\breve{\Upsilon}_5(x)$ are defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x)$ and $\Upsilon_5(x)$ in (4.7), but with $\breve{\kappa}_i$ in place of $\kappa_i$ where $\breve{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4$, $\breve{\kappa}_5 \geq \kappa_5$, and $\kappa_2 < \kappa_4$. We can verify that the assumptions of Theorem S.2 hold, as follows. Condition $(i)$ holds since $Av_1 = (1, -1, -1)^T$, $Av_2 = (0, 1, 1)^T$, $Av_3 = (1, 0, -1)^T$, $Av_4 = (0, 0, 1)^T$ and $Av_5 = (0, 1, 0)^T$. Assumption S.1 holds with $G^1 = \{1, 3\}$, $G^2 = \{2, 4\}$,

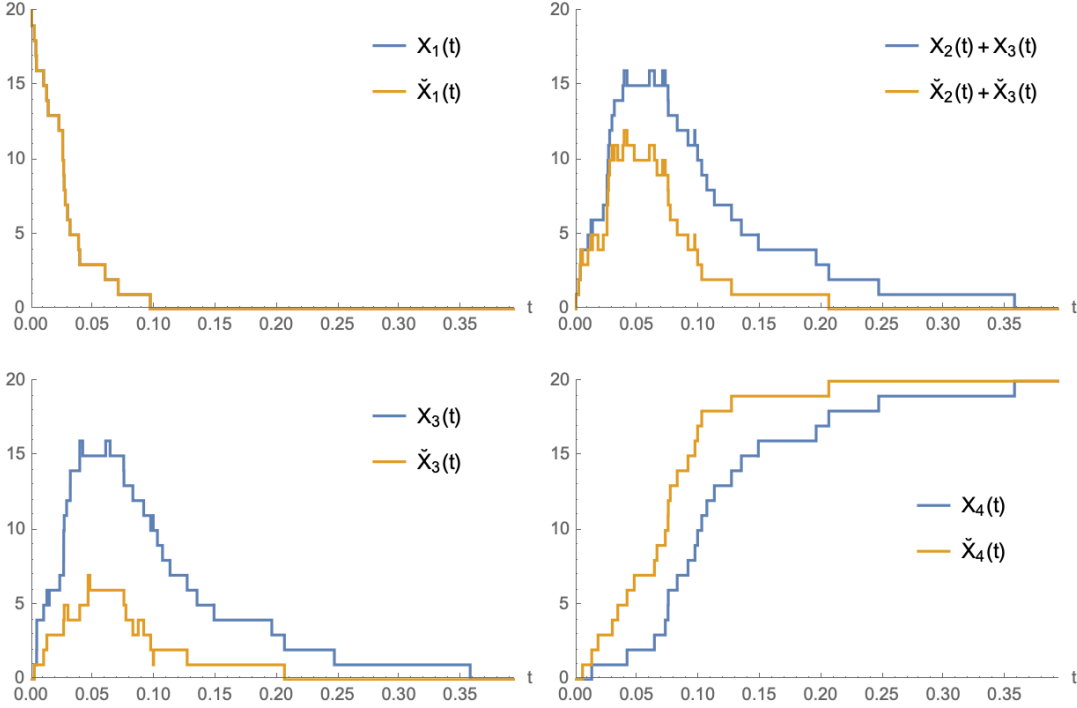Figure S.1: **A typical coupled realization of sample paths for Example 4.3.** Here $S_{tot} = 20$, $\kappa_1 = \check{\kappa}_1 = 30$, $\kappa_2 = \check{\kappa}_2 = 50$, $\kappa_3 = \check{\kappa}_3 = 10$, $\kappa_4 = \check{\kappa}_4 = 10$, $\kappa_5 = 1000$ and $\check{\kappa}_5 = 10$. Both processes $X$ and $\check{X}$ start at $(S_{tot}, 0, 0, 0)$. As shown in Section S.3.2, for these parameters, we have, almost surely, $X(t) \preceq_A \check{X}(t)$ for every $t \geq 0$ where the matrix $A$ is given in (S.21). A coupled realization using the algorithm described in Section S.4 is plotted to illustrate this result. In particular, we see in this sample that $X_1(t) \geq \check{X}_1(t)$, $X_2(t) + X_3(t) \geq \check{X}_2(t) + \check{X}_3(t)$ and $X_3(t) \geq \check{X}_3(t)$ for all times $t$. Moreover, the first passage time to $(0, 0, 0, S_{tot})$ for $X$ (which is equivalent to the first time to get to the state where $X_4 = S_{tot}$) is larger than for $\check{X}$. Since this first passage time result is true for all coupled samples, we can conclude that the mean first passage time from $(S_{tot}, 0, 0, 0)$ to $(0, 0, 0, S_{tot})$ is larger for $X$ than for $\check{X}$.

$G^3 = \{5\}$ and $\sigma(1) = 1$, $\sigma(2) = 3$, $\sigma(3) = 2$, $\sigma(4) = 4$, $\sigma(5) = 5$. To verify that condition (ii) of Theorem S.2 holds, let $x \in \mathcal{X}$, and first consider $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \geq w_2, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \leq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$. Since $y \in \partial_1(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 = \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = \kappa_3 x_1 = \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$, and so the desired inequality holds with equality. Secondly, consider $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 = w_2, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{2\bullet}, v_2 \rangle = \langle A_{2\bullet}, v_5 \rangle = 1$ and $\langle A_{2\bullet}, v_1 \rangle = -1$, we need to check that $\Upsilon_1(x) \geq \check{\Upsilon}_1(y)$, $\Upsilon_2(x) \leq \check{\Upsilon}_2(y)$ and $\Upsilon_5(x) \leq \check{\Upsilon}_5(y)$. Since $y \in \partial_2(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 \geq \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$, $\Upsilon_2(x) = \kappa_2 x_2 = \kappa_2 y_2 = \check{\kappa}_2 y_2 = \check{\Upsilon}_2(y)$ and $\Upsilon_5(x) = \kappa_5 x_2 = \kappa_5 y_2 \leq \check{\kappa}_5 y_2 = \check{\Upsilon}_5(y)$, and so the desired inequality holds. Lastly, consider $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 \geq w_2, x_2 + x_3 = w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = -1$ and $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_4 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \geq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$ and $\Upsilon_2(x) + \Upsilon_4(x) \leq \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$. For $y \in \partial_3(K_A + x) \cap \mathcal{X}$, since $\kappa_2 < \kappa_4$, we have that $\Upsilon_1(x) + \Upsilon_3(x) = (\kappa_1 + \kappa_3) x_1 \geq (\kappa_1 + \kappa_3) y_1 = (\check{\kappa}_1 + \check{\kappa}_3) y_1 = \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$ and

10

$\Upsilon_2(x) + \Upsilon_4(x) = \kappa_2 x_2 + \kappa_4 x_3 = (\kappa_4 - \kappa_2)x_3 + \kappa_2(x_2 + x_3) \leq (\kappa_4 - \kappa_2)y_3 + \kappa_2(y_2 + y_3) = \check{\kappa}_2 y_2 + \check{\kappa}_4 y_3 = \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$. Thus, the conditions of Theorem S.2 are satisfied. In particular, we can conclude when $\kappa_2 < \kappa_4$, that it takes less time to get to $(0,0,0,\mathrm{S_{tot}})$ from $(\mathrm{S_{tot}}, 0, 0, 0)$ if reaction ⑤ is added to the system without that reaction[9].

## S.4   An algorithm for coupled stochastic simulation

We now provide an algorithm for stochastic simulation of the coupled continuous-time Markov chains $X$ and $\check{X}$ under the conditions of Theorems 3.1, 3.2, 3.3 or S.2, when the transitions rates are bounded on the state space, i.e., when (5.1) holds.

---

**Algorithm:** Stochastic simulation for coupled continuous-time Markov chains $X$ and $\check{X}$.

**Data:** Integer $n \geq 1$, real $T > 0$, set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$, $x^\circ, \check{x}^\circ$ in $\mathcal{X}$, functions $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \ldots, \check{\Upsilon}_n)$ and integer $a \in \{1, 2, 3, 4\}$ to indicate which theorem is invoked (3.1, 3.2, 3.3 or S.2).

**Result:** Sample of initial time and subsequent potential jump times $T_0, T_1, \ldots, T_N$ and associated states $X(T_0), X(T_1), \ldots, X(T_N)$ and $\check{X}(T_0), \check{X}(T_1),$ $\ldots, \check{X}(T_N)$ for the continuous-time Markov chains $X$ and $\check{X}$ in the time interval $[0, T]$.

$\lambda \leftarrow 1 + n \max \left\{ \sup_{x \in \mathcal{X}} \sum_{j=1}^n \Upsilon_j(x), \sup_{x \in \mathcal{X}} \sum_{j=1}^n \check{\Upsilon}_j(x) \right\}$ ;

$K \leftarrow 0$;

$T_0 \leftarrow 0$;

**while** $T_K \leq T$ **do**

    $T_{K+1} \leftarrow T_K +$ `Exponential`$(\lambda)$;

    $K \leftarrow K + 1$;

**end**

$Y_0, \check{Y}_0 \leftarrow x^\circ, \check{x}^\circ$;

$X(T_0), \check{X}(T_0) \leftarrow x^\circ, \check{x}^\circ$;

$N \leftarrow K - 1$;

**if** $N \geq 1$ **then**

    **for** $k \leftarrow 0$ **to** $N - 1$ **do**

        $U \leftarrow$ `Uniform`$([0, 1])$;

        $Y_{k+1}, \check{Y}_{k+1} \leftarrow$ `TransitionDTMC`$(\Upsilon, \lambda, Y_k, U, a)$,

        `TransitionDTMC`$(\check{\Upsilon}, \lambda, \check{Y}_k, U, a)$;

        $X(T_{k+1}), \check{X}(T_{k+1}) \leftarrow Y_{k+1}, \check{Y}_{k+1}$;

    **end**

**end**

---

The random variables $T_1, \ldots, T_N$ are called potential jump times because it could be that $X(T_k) = X(T_{k+1})$ or $\check{X}(T_k) = \check{X}(T_{k+1})$ for some $0 \leq k \leq N - 1$. Letting $T_{N+1} := T$, the trajectories of $X$ are given by $X(t) = X(T_k)$ for $T_k \leq t < T_{k+1}$, $0 \leq k \leq N$, and similarly for the trajectories of $\check{X}$.

---

[9]The system without reaction ⑤ can be obtained from the system with reaction ⑤ by setting $\kappa_5 = 0$. While strictly speaking a zero rate constant is not within our definition of mass action kinetics, our theory does cover propensity functions with such a zero rate constant.

The function `TransitionDTMC` can be found below. This function is meant to replicate $\Phi_\lambda(x, u)$ in (5.4) for the case of Theorems 3.1 and 3.2, and $\Psi_\lambda(x, u)$ in (5.34) for the case of Theorems 3.3 and S.2.

---

**Function** `TransitionDTMC`$(\Upsilon, \lambda, x, u, a)$

    **Data:** Integer $n \geq 1$, set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, vectors $v_1, \ldots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$.

    **Input:** Function $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$, $\lambda > 0$, $x \in \mathcal{X}$, $u \in [0, 1]$. Integer
            $a \in \{1, 2, 3, 4\}$ to indicate which theorem is invoked (3.1, 3.2, 3.3 or S.2).
            For the case of Theorem 3.3 or S.2, include partition $\{G^1, \ldots, G^s\}$ and
            bijection $\sigma$.

    **Output:** State $x + v \in \mathcal{X}$.

    $v \leftarrow 0$;

    **if** $a \in \{1, 2\}$ **then** // The case of Theorem 3.1 or 3.2.

        **for** $j \leftarrow 1$ **to** $n$ **do**

            **if** $\frac{j-1}{n} \leq u < \frac{j-1}{n} + \frac{\Upsilon_j(x)}{\lambda}$ **then**

                $v \leftarrow v_j$;

            **end**

        **end**

    **end**

    **else** // The case of Theorem 3.3 or S.2.

        $p_0 \leftarrow 0$;

        **for** $k \leftarrow 1$ **to** $s$ **do**

            $p_k \leftarrow p_{k-1} + |G^k|$;

            **for** $q \leftarrow p_{k-1} + 1$ **to** $p_k$ **do**

                **if** $\frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^{q-1} \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda} \leq u < \frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^{q} \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda}$ **then**

                    $v \leftarrow v_{\sigma(q)}$;

                **end**

            **end**

        **end**

    **end**

    **return** $x+v$

---

**Remark S.3.** The above algorithm can be adapted to provide simultaneous stochastic simulation for $X$ and $\check{X}$ when transition rates are not bounded on the state space, by applying the algorithm on a sequence of bounded sets, which expand to the whole state space. This employs a sequence of successively defined stopping times $\tau_0 = 0$, $\tau_\ell = \inf\{t \geq \tau_{\ell-1} \mid X(t) \notin C_\ell \text{ or } \check{X}(t) \notin C_\ell\}$, $\ell = 1, 2, \ldots$, where the $C_\ell$ are compact, $C_\ell \subseteq C_{\ell+1}$ for $\ell = 1, 2, \ldots$ and $\cup_{\ell=1}^\infty C_\ell = \mathcal{X}$. The simulation of the pair $(X, \check{X})(t)$ for $\tau_\ell \leq t < \tau_{\ell+1}$ uses the above algorithm on $C_\ell$ for $\ell = 1, 2, \ldots$.

# References

[1] MEYN, S. P., AND TWEEDIE, R. L.(1993). Stability of Markovian processes iii: Foster–Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3):518—548.

[2] SHARPE, M. (1988). *General Theory of Markov Processes.* Academic Press.