Constant order multiscale reduction for stochastic reaction networks

German Enciso∗, Jinsu Kim†

October 14, 2019

Abstract

In this paper, we provide the accuracy of the multiscale model reduction for a stochastic reaction network under a certain multiscaling regime. When a subset of chemical species in a reaction network has much higher abundance than others, the associated stochastic process can be modeled under multiscaling regime. Previous works in the literature showed that the original stochastic system can be approximated with a reduced limiting system, as the scaling parameter $N$ tends to infinity. In general scaling regimes, the convergence to the limiting system can be studied by using the law of large numbers and the relative compactness in a measure space. In this work, by using the Kolmogorov equation we study the convergence rate of a stochastic network system in the multiscaling nature to a reduce system assuming that each reaction has a constant order propensity. We also provide the convergence of the marginal probability distribution.

1 Introduction

Regarding the intrinsic noise of a reaction network system, the copy number of the chemical species in the system can be modeled by a continuous-time, discrete-space Markov process. One of the typical goals for a stochastically modeled reaction network is to compute or to estimate the time evolution of the distribution. The distribution can be computed by solving the Kolmogorov equation, but this direct approach is rarely valid because of its high dimensionality. Therefore computational and analytic methods for estimating the distribution associated with a stochastic reaction network system are developed [12, 18, 19, 6, 7, 14, 5, 13, 15, 16].

In a stochastic system with multiscaling setup, some chemical species in a network system have much higher molecular counts that the others. Furthermore, the rate constant assigned to each reaction can also have different orders of magnitude. Hence each species can be modeled with different time scale due to the different orders of reaction propensities under usual mass action kinetics [14]. In this case, the limiting behavior of the associated multiscale stochastic system could be either a stochastic, deterministic or so-called hybrid system. [14, 2, 5, 13, 9].

∗Department of Mathematics, University of California, Irvine, USA. enciso@uci.edu
†Department of Mathematics, University of California, Irvine, USA. jinsu.kim@uci.edu
One of the special cases is that each reaction propensity is of constant order. Under this scaling regime, all species are slow. Under this circumstance relatively small amount of fluctuations for the high abundance species could be ignored up to a finite time, so that they are even assumed to be fixed at their initial counts \([10]\).

In this paper, using the same scaling regime, we provide the convergence rate for the distribution \(p^N(t)\) of the multiscale stochastic network system to the distribution \(p(t)\) of the limiting network system, as \(N \to \infty\) for a fixed time \(t\). We also provide a stronger convergence result by showing the convergence of the distribution of individual species (i.e. the marginal distribution of \(p^N\)).

This manuscript has the following outline. In Section 2 we state the basic notions of a stochastic system for a reaction network. In the same section we also provide multiscale setting for the stochastic model and the idea of network projection. The main theorem of this paper and the proof would be introduced in Section 3. Some propositions and necessary lemmas for the main results are given and proved in Appendices.

\section{Preliminaries}

\subsection{stochastic reaction network systems}

In this section, we provide a stochastic model associated with a chemical reaction network. We first introduce reaction networks. A chemical interacting system can be described with a reaction network. A reaction network is a graphical configuration consisting of constituent species, complexes that are combinations of species, and reactions between complexes. A triple \((S, C, R)\) represents a reaction network where \(S, C\) and \(R\) are collections of species, complexes and reactions, respectively. We formally define a reaction network \((S, C, R)\).

\textbf{Definition 2.1.} A reaction network is defined with a triple of finite sets \((S, C, R)\) such that

1. the species set \(S = \{S_1, S_2, \ldots, S_d\}\) contains the species of the reaction network,

2. the complex set \(C = \{y_1, y_2, \ldots, y_c\}\) contains complexes \(y_k\)'s where for each \(k\),

\[ y_k = \sum_{i=1}^{d} y_{ki}S_i \quad \text{for some non-negative integers } y_{ki}, \text{ and} \]

3. the reaction set \(R = \{R_1, R_2, \ldots, R_r\}\) consists of ordered pairs \((y, y')\) such that \(y, y' \in C\).

In the graphical configuration of a reaction network \((S, C, R)\), we use \(y \to y'\) with \(y, y' \in C\) to denote a reaction \((y, y') \in R\). For defining a dynamical system associated with a reaction network \((S, C, R)\), we denote a complex \(y_k\) by either a combination of species such as \(\sum_{i=1}^{d} y_{ki}S_i\) or a \(d\)-dimensional vector \(y_k = (y_{k1}, \ldots, y_{kd})^T\), interchangeably. In case \(y_{ki} = 0\) for all \(i\), the corresponding complex \(y_k = \sum_{i=1}^{d} y_{ki}S_i\) is denoted by \(\emptyset\) in the graphical configuration of the reaction network.
Example 2.1. Consider the following reaction network describing a substrate-enzyme system with a protein dilution.

\[ S + E \rightleftharpoons SE \rightarrow E + P, \quad P \rightarrow 0 \]

For this reaction network, \( \mathcal{S} = \{ S, E, SE, P \} \), \( \mathcal{C} = \{ S + E, SE, E + P, 0 \} \) and \( \mathcal{R} = \{ S + E \rightarrow SE, SE \rightarrow S + E, SE \rightarrow E + P, P \rightarrow 0 \} \).

For the usual stochastic model, each component of the \( d \)-dimensional continuous time, discrete state Markov process \( X(t) = (S_1(t), \ldots, S_d(t)) \in \mathbb{Z}_d^{\geq 0} := \{ x \in \mathbb{Z}^d : x_i \geq 0, i = 1, 2, \ldots, d \} \) gives the counts of each species in a reaction network at time \( t \). The transitions of the Markov process are determined by the reactions. In order to define the transition probabilities, we use state-dependent intensity (or propensity) functions \( \lambda_k : \mathbb{Z}_d^{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \) of reaction \( y_k \rightarrow y_k' \). Then we assume that

\[
P(X(t + \Delta t) = z + y' - y \mid X(t) = z) = \sum_{y_k \rightarrow y_k' \in \mathcal{R}} \lambda_k(z)\Delta t + o(\Delta t),
\]

for each state \( z \) in the state space \( \mathcal{S} \) of the associated Markov process \( X \). In this paper, the counts of a species \( S \) at time \( t \) will be denoted by \( S(t) \) and \( z \) will represent a state in the state space \( \mathcal{S} \) of the associate Markov process \( X \).

Let \( E_x \) denote the expectation of \( X \) with \( X(0) = x \) and let \( p_x(A, t) = P(X(t) \in A \mid X(0) = x) \) for \( A \subset \mathbb{Z}_d^{\geq 0} \) and \( x \in \mathbb{Z}_d^{\geq 0} \). We often omit the subindex \( x \) when the circumstance does not incorporate an initial condition \( x \). Then for a non-explosive Markov process \( X \) associated to a reaction network \( (\mathcal{S}, \mathcal{C}, \mathcal{R}) \), each \( p(z, t) \) solves Kolmogorov forward equation (master equation)

\[
\frac{d}{dt} p(z, t) = \sum_k \lambda_k(z - y_k' + y_k)p(z - y_k' + y_k, t) - \sum_k \lambda_k(z)p(z, t), \tag{2}
\]

where \( \sum_k \) denotes the sum over all reactions in \( \mathcal{R} \). The usual choice of intensity of a reaction \( y_k \rightarrow y_k' \) in a network system \( (\mathcal{S}, \mathcal{C}, \mathcal{R}, \mathcal{K}) \) with a set of rate constants \( \mathcal{K} \) is

\[
\lambda_k(x) = \kappa_k x^{(y)}, \tag{3}
\]

where \( u^{(v)} = \prod_{i=1}^d \frac{u_i^{d}}{(u_i - v_i)!} 1_{\{u_i \geq v_i\}} \) for \( u, v \in \mathbb{Z}_d^{\geq 0} \). This choice of intensities is termed stochastic mass action kinetics. The stochastic process \( X(t) \) also has another representation, so-called random time change representation. \[17\]

\[
X(t) = X(0) + \sum_k Y_k \left( \int_0^t \lambda_k(X(s)) ds \right) (y_k' - y_k), \tag{4}
\]

where \( Y_k \)’s are independent unit Poisson random variables.

In the associated graph of a reaction network, the rate constants typically appeared next to the reaction arrow as in \( y \rightarrow y' \). Letting \( \mathcal{K} \) be a set of rate constants, \( (\mathcal{S}, \mathcal{C}, \mathcal{R}, \mathcal{K}) \) stands for a system dynamics associated with a reaction network \( (\mathcal{S}, \mathcal{C}, \mathcal{R}) \). We call \( X(t) \) with the probabilities (1) a stochastic process associated with a network system \( (\mathcal{S}, \mathcal{C}, \mathcal{R}, \mathcal{K}) \). Throughout this manuscript, we model a stochastic process with the stochastic mass action kinetic (3) for a reaction network.
2.2 Multiscaling for reaction networks

In this section, we provide a multiscale stochastic model associated to a reaction network. We use the same multiscaling model introduced in [14, 5, 13]. In the entire paper, we use the conventional ‘big O’ notation. Let $N$ be a scaling parameter which could be interpreted as either the volume of the system, Avogadro’s number or any biological parameter.

Let $X^N(t) = (S^N_1(t), \ldots, S^N_d(t))$ be a stochastic model associated to $(S, C, R, K)$ for which each species has different order of abundance. Typically it has two sets of scaling exponents, $\{\alpha_i | S_i \in S\}$ and $\{\beta_k | y_k \rightarrow y'_k \in R\}$. Each $\alpha_i$ represents the size of the abundance of species $S_i$ such that

$$N^{-\alpha_i}S^N_i(0) = O(1). \quad (5)$$

If $\alpha_i = 1$, then $N^{-\alpha_i}S^N_i(t)$ may represent the concentration of $S_i$ at time $t$.

We also assume the rate constants of each reaction have different order of magnitude so we define scaled intensities $\lambda^N_k$’s as

$$\lambda^N_k(x) = N^{\beta_k} \lambda_k(x). \quad (6)$$

That is, a rate constant $\kappa_k$ is scaled as $N^{\beta_k} \kappa_k$ for each reaction $y_k \rightarrow y'_k$. For a given set of rate constants $K = \{\kappa_k : y_k \rightarrow y'_k \in R\}$, we denote the set of the scaled rate constants (6) by $K^N$. Then by the representation (4), the scaled process $Z^N_i(t) = N^{-\alpha_i}S^N_i(t)$ for each $i = 1, 2, \ldots, d$ solves

$$Z^N_i(t) = Z^N_i(0) + \sum_k N^{-\alpha_i} Y_k \left( \int_0^t N^{\beta_k+y_k-\alpha_i} \lambda_k(Z^N(s))ds \right) (y'_k - y_k), \quad (7)$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$. Under time scaling $Z^N_{i,\gamma} = Z^N(N^\gamma t)$, we have

$$Z^N_{i,\gamma}(t) = Z^N_{i,\gamma}(0) + \sum_k N^{-\alpha_i} Y_k \left( \int_0^t N^{\gamma+y_k-\alpha_i} \lambda_k(Z^N_{i,\gamma}(s))ds \right) (y'_k - y_k). \quad (8)$$

When $z$ is not an integer valued vector, we read $\lambda_k(z)$ as

$$\lambda_k(z) = \kappa_k \frac{1}{N^{\alpha_y}} z^{(y_k)} = \kappa_k \prod_{i=1}^d z_i(z_i - N^{-\alpha_i}) \cdots (z_i - (y_k - 1) N^{-\alpha_i}).$$

One can find more details about the backgrounds of this multiscale setting in [14, 5, 13].

2.3 Projection of reaction networks

In this section, we introduce the projection of reaction networks. We begin with an example. Consider the following reaction network,

$$A + B \xrightarrow{a/N} 2B. \quad (9)$$
Suppose the associated Markov process \(X(t) = (X_A(t), X_B(t))\) started at \((X_A(0), X_B(0))\) where \(X_A(0) = 10\) and \(X_B(0) = N\), with a scaling parameter \(N\). The stochastic mass action (3) defines the intensity of the reaction \(A + B \rightarrow 2B\) at time \(t\) as
\[
\lambda_{A+B \rightarrow 2B}(X(t)) = \frac{a}{N} X_A(t) X_B(t).
\]
Until \(t < \infty\), we expect the reaction fires finitely many times and hence \(\frac{a}{N} X_A(t) X_B(t) \rightarrow aX_A(t)\), as \(N \rightarrow \infty\). This means that if \(N\) is sufficiently large, the counts of species \(A\) in this network behaves similarly as the counts of species \(A\) in the following network,

\[A \xrightarrow{a} \emptyset.\] (10)

Network (10) is a projection of network (9) in the sense that we only count species \(A\) assuming \(B\) is fixed at \(N\). As this example shows, a projected network can be used to describe an asymptotic behavior of certain species under a scaling regime for the original network system.

We first define two projection functions for complexes and reactions in \((\mathcal{S}, \mathcal{C}, \mathcal{R})\) with \(\mathcal{S} = \mathcal{S}_L \cup \mathcal{S}_H\) where \(\mathcal{S}_L = \{S_1, S_2, \ldots, S_d\}\) and \(\mathcal{S}_H = \{S_{d+1}, S_{d+2}, \ldots, S_{d+r}\}\). Let \(q_L : \mathbb{Z}^{d+r} \rightarrow \mathbb{Z}^d\) and \(q_H : \mathbb{Z}^{d+r} \rightarrow \mathbb{Z}^r\) be projection functions such that for each \(v = (v_1, \ldots, v_d, v_{d+1}, \ldots, v_{d+r})^T \in \mathbb{Z}^{d+r}\),

\[
q_L(v) = (v_1, v_2, \ldots, v_d)^T \in \mathbb{Z}^d \quad \text{and} \quad q_H(v) = (v_{d+1}, v_{d+2}, \ldots, v_{d+r})^T \in \mathbb{Z}^r,
\]

respectively.

We demonstrate the usage of \(q_L\) and \(q_H\) with the network (9) for which we set \(\mathcal{S}_L = \{A\}\) and \(\mathcal{S}_H = \{B\}\). Since the associated vector for the complex \(y = A + B\) in (9) is \(y = (1, 1)^T\), we have \(q_L(y) = 1\) which means one \(A \in \mathcal{S}_L\). Similarly, we have \(q_H(y) = 1\) meaning that one \(B \in \mathcal{S}_H\). As a complex is denoted by a combination of species, we also let \(q_L(A + B) = A\) and \(q_H(A + B) = B\). In the same way, for \(y \rightarrow y'\) where \(y = A + B\) and \(y' = 2B\), the projected reaction by \(q_L\) is \(q_L(A + B) \rightarrow q_L(2B)\) where \(q_L(y) = A\) and \(q_L(y') = B\).

In this way, the projected network \((\mathcal{S}_L, \mathcal{C}_L, \mathcal{R}_L)\) of the original reaction network \((\mathcal{S}, \mathcal{C}, \mathcal{R})\) by \(q_L\) is defined with

\[
\mathcal{S}_L = \{S_1, \ldots, S_d\}, \quad \mathcal{C}_L = \{q_L(y) : y \in \mathcal{C}\}, \quad \text{and}
\]

\[
\mathcal{R}_L = \{q_L(y) \rightarrow q_L(y') : y \rightarrow y' \in \mathcal{R} \text{ such that } q_L(y') - q_L(y) \neq \vec{0}\}.\] (12)

If the set of rate constants \(\mathcal{K}\) are given to the original network \((\mathcal{S}, \mathcal{C}, \mathcal{R})\), then the rate constants for the reduced system \((\mathcal{S}_L, \mathcal{C}_L, \mathcal{R}_L)\) are inherited as

\[
\mathcal{K}_L = \left\{ \tilde{\kappa}_\ell = \sum_k \kappa_k : \tilde{y}_\ell \rightarrow \tilde{y}'_\ell \in \mathcal{R}_L \right\}.\] (13)

The stochastic process associated with the projected network system \((\mathcal{S}_L, \mathcal{C}_L, \mathcal{R}_L, \mathcal{K}_L)\) is also modeled with the stochastic mass-action kinetics.

Note that some species in \(\mathcal{S}_L\) could be negligible because it neither appears in any reactions in \(\mathcal{R}_L\) nor changes its count by any reaction in \(\mathcal{R}_L\). For example, if \(\mathcal{S}_L = \{A, B\}\), \(\mathcal{S}_H = \{C\}\) and
\( \mathcal{R} = \{ A + C \rightarrow C, A + B \rightarrow A \} \), then the \( \mathcal{R}_L \) consists of a single reaction \( A + B \rightarrow A \). In this projected network, the copy number of \( A \in \mathcal{S}_L \) never changes. For the sake of simplicity, we do not consider such a case.

As the rate constant in the example (9) is scaled with \( N \), the projection we use in this paper is based on multiscale among the counts of species. This multiscaling regime is introduced in [10]. The author of the paper reduced a given network by freezing the counts of species whose initial copy number is of order \( N \). Under the same principle, we project an original network on to the space of species of low copy numbers obtained by freezing species of high copy numbers. Thus we term \( (\mathcal{S}_L, \mathcal{C}_L, \mathcal{R}_L) \) the projected network by freezing the species in \( \mathcal{S}_H \).

Example 2.2. Let the following network be \( (\mathcal{S}, \mathcal{C}, \mathcal{R}) \).

\[
A + B \xrightarrow{\kappa_1} C, \quad A + D \xrightarrow{\kappa_3} 2D, \quad D \xrightarrow{\kappa_5} A
\]

Let \( \mathcal{S} = \mathcal{S}_L \cup \mathcal{S}_H \) with \( \mathcal{S}_L = \{ A, B, C \} \) and \( \mathcal{S}_H = \{ D \} \). For \( (\mathcal{S}, \mathcal{C}, \mathcal{R}, \mathcal{K}) \), we generate the projected network system \( (\mathcal{S}_L, \mathcal{C}_L, \mathcal{R}_L, \mathcal{K}_L) \) of by freezing \( D \). We have \( q_L(A + B) = A + B, q_L(C) = C, q_L(A + D) = A, q_L(2D) = \emptyset, q_L(D) = \emptyset \) and \( q_L(A) = A \). Hence the projected network system \( (\mathcal{S}_L, \mathcal{C}_L, \mathcal{R}_L, \mathcal{K}_L) \) is

\[
A + B \xrightarrow{\kappa_1} C, \quad A \xrightarrow{\kappa_3 + \kappa_4} \emptyset.
\]

3 Main results

In this section, we introduce our main results. In [5, 14], it was shown that if scaling exponents \( \alpha_i \)'s and \( \beta_k \)'s and \( \gamma \) in (8) satisfy some balance conditions, then species of high abundance follow a system of differential equations with random coefficients, and the species of low abundance follow a piece-wise deterministic Markov process. In this paper, we assume that \( \alpha_k \)'s, \( \beta_k \)'s and \( \gamma \) are chosen so that the species in \( \mathcal{S}_H \) are confined around their initial values with high probability until a finite time \( t \). Under this scaling regime, we show that the species in \( \mathcal{S}_L \) approximately follow the projected network system \( (\mathcal{S}_L, \mathcal{C}_L, \mathcal{R}_L, \mathcal{K}_L) \) until \( t \) as the scaling parameter tends to infinity.

This somewhat special scaling regime allows us to investigate the convergence of marginal distributions as well as the convergence rate. These extra outcomes may not be guaranteed with the usual weak convergence of probability measures. In order to investigate such results, we assume that the projected system admits a stationary distribution (a solution of the Kolmogorov’s equation (2)) with a fast-decaying tail. Then it follows that not only the marginal distribution of the original system converges to the corresponding marginal of the projected network system, but also that the convergence rate can be evaluated for a fixed \( t \) with respect to \( N \).

3.1 Lemmas for the main theorem

We first give a list of the notations we use throughout this manuscript. Proofs of some lemmas are introduced in Appendix A

1. For the proof of the necessary lemmas and the main theorem, we use abbreviated notations \( Z^N \) and \( Z \) instead of \( Z^{N, \gamma} \) and \( Z^\gamma \), respectively.
2. Let $\mathcal{K}$ be a given set of rate parameters for $(S, C, \mathcal{R})$. For each $N$, let $Z^N(t) = (Z_1^N(t), \ldots, Z_{d+r}^N(t))$ be the stochastic process associated with $(S, C, \mathcal{R}, \mathcal{K}^N)$ as introduced in Section 2.2, where $\mathcal{K}^N$ is the set of scaled rate constants as (6). Let $Z(t)$ be the stochastic process associated with $(S_L, C_L, \mathcal{R}_L, \mathcal{K}_L)$, where $\mathcal{K}_L$ is as (13). We define

$$\kappa = \sum_{y_k \rightarrow y'_k \in \mathcal{R}} \kappa_k.$$  \hspace{1cm} (15)

3. In order to distinguish the intensities for $(S_L, C_L, \mathcal{R}_L, \mathcal{K}_L)$ from the intensities for $(S, C, \mathcal{R}, \mathcal{K})$, we use $\lambda_{L,u}$ for the intensity of each reaction $\bar{y}_u \rightarrow \bar{y}'_u \in \mathcal{R}_L$. The intensity function of a reaction $y_k \rightarrow y'_k \in \mathcal{R}$ can be split into two parts as

$$\lambda^N_k(z) = \lambda^N_{L,k}(z)\lambda^N_{H,k}(z),$$  \hspace{1cm} (16)

where $\lambda^N_{L,k}(z) = \kappa_k z_{l_k}(q_{l_k}(y_k))$ and $\lambda^N_{H,k}(z) = z_{l_k}(q_{H}(y_k))$. Then by the definition of the projection and from (8), the intensity of $Z$ associated with a reaction $\bar{y}_u \rightarrow \bar{y}'_u \in \mathcal{R}_L$ is

$$\lambda_{L,u}(z) = \sum_{y_i \rightarrow y'_i \in \mathcal{R}} \sum_{q_{l_k}(y_k)=y_i,q_{H}(y_k)=y'_i} N^{\gamma+\beta_i+\gamma_i-\alpha_i}h_{L,i}(z).$$  \hspace{1cm} (17)

4. We denote the distribution of $Z^N(t)$ by $p^N(\cdot,t)$. In particular, for $A \subset Z_{d\geq0}$ and $A' \subset Z_{r\geq0}$,

$$p^N(S,S',t) = P\{ (Z_1^N(t), \ldots, Z_d^N(t)) \in A \cap (Z_{d+1}^N(t), \ldots, Z_{d+r}^N(t)) \in A' \}.$$

The distribution of the stochastic process $Z(t)$ is denoted by $p(\cdot,t)$.

5. We use the usual $\infty$-norm and 1-norm for vectors in $Z^m$. That is, for $v \in Z^m$

$$\|v\|_\infty = \max_i |v_i| \quad \text{and} \quad \|v\|_1 = \sum_{i=1}^m |v_i|.$$

Then we let

$$c_L = \max_{y_k \rightarrow y'_k \in \mathcal{R}} \{\|q_L(y_k)\|_1\} \quad \text{and} \quad c_H = \max_{y_k \rightarrow y'_k \in \mathcal{R}} \{\|q_H(y_k)\|_1\}.$$  \hspace{1cm} (18)

6. Note that the state space of $Z^N$ is a subset of $Z_{d\geq0} \times \mathbb{R}_{r\geq0}^d$. Note further that the state space of $(Z_{d+1}^N, \ldots, Z_{d+r}^N)$ depends on the scaling parameter $\gamma$, as the scaling exponent $\alpha_i$ in (8) is 1 for $i = d+1, \ldots, d+r$.

We denote the state space of $Z^N$ by $S_{\gamma} \times S_{\gamma}$, where $S_{\gamma} \subset Z_{d\geq0}$ and $S_{\gamma} \subset \mathbb{R}_{r\geq0}^d$. Let $Z^N(0) = x = (x_{\ell}, x_h)$ be the initial condition of $Z^N$ such that $q_L(x) = x_{\ell}$ and $q_H(x) = x_h$. We define $S_{L,M} = \{z_{\ell} \in S_{\gamma} : \|z_{\ell}\|_\infty \leq M\}$ and $S_{H,M} = \{z_h \in S_{\gamma} : \|z_h - x_h\|_\infty \leq M\}$ for some $M > 0$. Let $S_M = S_{L,M} \times S_{H,M}$.

7. The transitions of the $i$-th component of the scaled process $Z^N_i$ are given by $(y'_{k,i} - y_{k,i})/N^{\alpha_i}$, where $y_k \rightarrow y'_k \in \mathcal{R}$. We denote this scaled reaction vector by $y_k^N - y_k^N$ whose $i$-th component is equal to $(y'_{k,i} - y_{k,i})/N^{\alpha_i}$.
The followings are assumptions for the main theorem.

(CD1) We suppose species in \((S, C, R)\) are classified into two classes; high and low copy number species. Thus for each species \(S_i \in S\), the scaling exponents \(\alpha_i\)'s shown in (5) are chosen to be

\[\alpha_i \in \{0, 1\}.\]

Let \(S_L = \{S_i : \alpha_i = 0\} = \{S_1, \ldots, S_d\}\), and let \(S_H = \{S_i : \alpha_i = 1\} = \{S_{d+1}, \ldots, S_{d+r}\}\).

(CD2) We assume that \(Z^N(0) = (x_t, x_h)\) for a \(x_t \in \mathbb{Z}_{\geq 0}^d\) and a \(x_h \in \mathbb{R}^r\). We further assume that for \(i = 1, 2, \ldots, r\),

\[S_{d+i}(0) = N^{\alpha_{d+i}}\] so that each component of \(x_h\) is 1.

(CD3) We choose the scaling exponents for reaction rate constants \(\beta_k\)'s shown in (6) and the time scaling exponent \(\gamma\) so that all reactions are fired evenly likely with respect to \(N\). Thus for each reaction \(y_k \rightarrow y'_k\), we assume

\[\gamma + \beta_k + y_k \cdot \alpha = 0.\]

This implies that all reaction intensities in (8) are of order 1. In this case, we use abbreviated notations \(Z^N\) and \(Z\) instead of \(Z^{N,\gamma}\) and \(Z^\gamma\), respectively. More general condition for \(\gamma\) that prevents the system from explosion is introduced in [14] being termed the species balance condition.

(CD4) We also assume that the stochastic system associated with the projected network \((S_L, C_L, R_L, K_L)\) admits a stationary distribution \(\pi\) such that for any positive integer \(m\)

\[\sum_{x \in \mathbb{Z}_+^d} |x_1 + \cdots + x_d|^m \pi(x) < \infty.\]

In [4], it is known that the stochastic system associated with complex balanced reaction networks, that arise in many biology systems, admits a product form of Poissons stationary distribution satisfying (CD4). Note that (CD2) can be generalized for the main theorem.

One of the key ideas for the main theorem in this paper is to show that \(Z^N\) stays in \(S_M\) until a finite time \(t\) with high probability. In the following lemmas we show that the probability of \(Z^N\) to escape \(S_M\) is low if \(N\) is sufficiently larger than \(M\).

Lemma 3.1. Suppose (CD1), (CD2) and (CD3) hold for the stochastic process \(Z^N\) associated with \((S, C, R, K^N)\). Let \((S_L, C_L, R_L, K_L)\) be the projection obtained by freezing species in \(S_H\). For a state \(z\) of the stochastic process \(Z^N\), we denote \(z = (z_t, z_h)\) such that \(q_L(z) = z_t \in \mathbb{Z}_+^d\) and \(q_H(z) = z_h \in \mathbb{R}_{\geq 0}^r\). Let \(y_k \rightarrow y'_k \in R\) and \(\bar{y}_u \rightarrow \bar{y}'_u \in R_L\) such that \(q_L(y_k) = \bar{y}_u\) and \(q_L(y_k) = \bar{y}'_u\). Then for each \(z \in S_M\), there exist \(c > 0\) such that

\[\kappa \left(1 - \frac{cM}{N}\right) \lambda_{L,k}^N(z_t) \leq \lambda_k^N(z) \leq \kappa \left(1 + \frac{cM}{N}\right) \lambda_{L,k}^N(z_t) \text{ for } N \text{ sufficiently large.}\]

In particular, for \(z_h \in S_{H,M}\),

\[1 - \frac{cM}{N} \leq \lambda_{H,k}^N(z_h) \leq 1 + \frac{cM}{N} \text{ for } N \text{ sufficiently large.}\]
Proof. Note that by (17) and (CD3), we have
\[
\lambda_{L,a}(z_\ell) \leq \kappa \lambda_{L,k}(z_\ell).
\] (19)

Note that if \( z_h \in S_{H,M} \), then \((1 - \frac{M}{N})^{cH} \leq z_h^{\ell(qH(y_k))} \leq (1 + \frac{M}{N})^{cH} \), where \( c_H \) is as (18). Therefore, for sufficiently large \( N \), the Taylor expansion of function \((1 + x)^m\) as \( x \to 0 \) implies that \((1 - \frac{cM}{N}) \leq z_h^{\ell(qH(y_k))} \leq (1 + \frac{cM}{N})\) for some \( c > 0 \). This combined with (19) completes the proof. \( \square \)

Lemma 3.2. Let \( Z \) be the stochastic process associated to a network system \((S_L, C_L, R_L, K_L)\) with \( Z(0) = x_\ell \). Suppose (CD4) holds. Let \( J(t) \) be the number of jumps of \( Z \) by time \( t \). Then there exists a constant \( c \) such that
\[
E(J(t)^m) \leq c t^{m+1}
\]

By using Lemma 3.1 and Lemma 3.2, we prove Lemma 3.3. This lemma will be used to prove Lemma 3.6 which is one of the key lemmas for the main theorem.

Lemma 3.3. Suppose (CD1), (CD2) and (CD3) hold for the stochastic process \( Z^N \) associated with \((S, C, R, K^N)\). Let \((S_L, C_L, R_L, K_L)\) be the projection obtained by freezing species in \( S_H \). Let \( P^N \) and \( P \) be the probability measures of \( Z^N \) and \( Z \), respectively, such that \( Z^N(0) = (x_\ell, x_h) \) and \( Z(0) = x_\ell \). Let further \( \tau_n^N \) and \( \tau_n \) be the time of \( n \)-th jump of \( Z^N \) and \( Z \), respectively. Let \( t \) be a fixed time and let \( m \) be an arbitrary positive integer. Then for \( N \) sufficiently large, there is a constant \( c > 0 \) such that
\[
P^N(\tau_M^N < t) \leq P(\tau_M < t/2) \leq \left(\frac{1 + \frac{cM}{N}}{1 - \frac{cM}{N}}\right)^M \frac{ct^m}{M^m}. \] (20)

In particular, if \( M = f(N) \) such that \( \lim_{N \to \infty} \frac{f(N)^2}{N} = 0 \), then
\[
P^N(\tau_M^N < t) \leq P(\tau_M < t/2) \leq \frac{c't^m}{M^m}. \] (21)

for some \( c' > 0 \).

Another key step for the main theorem is to truncate the state space of \( Z^N \) and \( Z \) onto \( S_M \) and \( S_{L,M} \), respectively. On the truncated state spaces, the probability distributions \( p^N(\cdot, t) \) and \( p(\cdot, t) \) can be represented with finite-dimensional vectors.

To truncate the state spaces, we use finite state projection that has been proposed by [18, 19]. For a given continuous time, discrete state Markov chain \( X \), the FSP model of \( X \) is a Markov chain confined on on a finite set \( \widetilde{S} \). The transition rates of the FSP model are defined as the followings.

i) The transition rates between states inside \( \widetilde{S} \) are same as \( X \). ii) All transitions outgoing from \( \widetilde{S} \) are merged into a transition to a single state \( z_0 \). The corresponding transition rates are summed up. iii) The transitions between states in \( \widetilde{S}^c \) set to be zero. We define the FSP model formally.

Definition 3.1. Let \( X \) be a continuous time, discrete state Markov process, and let \( S \) be the state space of \( X \). A finite state projection (FSP) model of \( X \) on a finite set \( \widetilde{S} \subset S \) is a Markov chain \( \widetilde{X} \) for which there is a single state \( z_0 \in S \setminus \widetilde{S} \) such that
1. for \( z \neq z_0 \), \( P(X(t + \Delta t) = z|X(t) = x) = \begin{cases} P(X(t + \Delta t) = z|X(t) = x) & \text{if } x, z \in \widetilde{S}, \\
0 & \text{otherwise}, \end{cases} \)

2. \( P(X(t + \Delta t) = z_0|X(t) = x) = \begin{cases} \sum_{w \in \mathbb{S} \setminus \mathbb{S}'} P(X(t + \Delta t) = w|X(t) = x) & \text{if } x \in \mathbb{S} \\
0 & \text{otherwise}. \end{cases} \)

Under the conditions we assumed above, two FSP models of \( Z^N \) and \( Z \) truncated are close. Let \( z_0 \) and \( q_L(z_0) \) be the absorbing states of the two FSP models as defined in Definition 3.1. For the simplicity of the proof of Lemma 3.5, we assume that \( z_0 \) is not accessible for \( Z^N \) from \( \mathbb{S} \) with a single transition. Note that under (CD1), the \( i \)-th component scaled reaction vector \( y'_k - y_k \) is \( y'_{k,i} - y_{k,i} \) if \( i \in \{1, 2, \ldots, d\} \) and is \( \{y'_{k,i} - y_{k,i} \} \) if \( i \in \{d + 1, d + 2, \ldots, d + r\} \).

Let \( p_{fsp}(t) \) and \( p_{fL}(t) \) be probability distribution of \( Z^N_{fL}(t) \) and \( Z_{fL} \), respectively. Since the state space of two FSP models are finite, the probability distributions can be represented finite-dimensional vector. Hence for the distance of two vectors, we use the following lemma.

**Lemma 3.4.** Suppose \( A \) is an \( n \times n \) matrix such that for any vector \( u_0 \in \mathbb{R}^n_{\geq 0} \) with \( ||u_0||_1 = 1 \), system of differential equation

\[
\begin{align*}
\frac{d}{dt}u(t) &= Au, \\
u(0) &= u_0,
\end{align*}
\]

admits a unique solution \( u(t) \in \mathbb{R}^n_{\geq 0} \) with \( ||u(t)||_1 = 1 \) for any \( t \). Let \( A_{max} \) be the maximum entry of \( A \). Suppose \( v \) satisfies \( \frac{d}{dt}v(t) \leq Av(t) + b \) for some \( b \in \mathbb{R}^n_{\geq 0} \). Then for each \( t \),

\[ v(t) \leq tb + t^2nA_{max}b. \]

Now we introduce two key lemmas for the main theorem in Section 3.

**Lemma 3.5.** Suppose (CD1), (CD2) and (CD3) hold for the stochastic process \( Z^N \) associated with \( (S, C, R, \lambda^N) \). Let \( \{S_L, C_L, R_L, \lambda_L\} \) be the projection obtained by freezing species in \( S_H \). Let \( Z^N_{fL} \) and \( Z_{fL} \) be the FSP models of \( Z^N \) and \( Z \) on \( S_M \) and \( S_L,M \), respectively. Then for fixed \( t \) and for any \( A \subset S_L,M \), there exists \( c \) such that

\[ |p^N_{fL}(A, S_H,M, t) - p_{fL}(A, t)| \leq c \frac{M^{c+1+2d}}{N} \quad \text{for } M, N \text{ sufficiently large.} \]

**Proof.** Let \( z_0 \) and \( q_L(z_0) \) be the absorbing states of \( Z^N_{fL} \) and \( Z_{fL} \) as defined in Definition 3.1, respectively. We denote \( z_{\ell,0} = q_L(z_0) \) and \( z_{h,0} = q_H(z_0) \). For each \( (z_{\ell}, z_{h}) \in (S_L,M, S_H,M) \), the distributions \( p^N_{fL} \) and \( p_{fL} \) satisfy the Kolmogorov forward equation (chemical master equation) (2).

\[
\frac{d}{dt}p^N_{fL}(z_{\ell}, z_{h}, t) = \sum_k \lambda^N_{L,k}(z_{\ell} - q_L(y^N_k - y^N_k))\lambda^N_{H,k}(z_{h} - q_H(y^N_k - y^N_k)) \\
\times p^N_{fL}(z_{\ell} - q_L(y^N_k - y^N_k), z_{h} - q_H(y^N_k - y^N_k), t) \\
- \sum_k \lambda^N_{L,k}(z_{\ell})\lambda^N_{H,k}(z_{h})p^N_{fL}(z_{\ell}, z_{h}, t),
\]

and

\[
\frac{d}{dt}p_{fL}(z_{\ell}, t) = \sum_u \lambda_{L,u}(z_{\ell} - \bar{y}_u + \bar{y}_u)p_{fL}(z_{\ell} - \bar{y}_u + \bar{y}_u, t) - \sum_u \lambda_{L,u}(z_{\ell})p_{fL}(z_{\ell}, t),
\]

10
respectively.

We will show that after summing (23) over \( z_h \in S_{H,M} \), two differential equations (23) and (24) are very close. In order to do that, we go through the following two steps.

1. We have also, for a \( z_\ell \in S_{L, M} \)

\[
\sum_{y_k \rightarrow y'_k \in \mathcal{R}} \lambda^N_{L,k}(z_\ell - q_L(y'_k - y_k)) \leq c_1\kappa M^{c_\ell} \quad \text{and} \quad \sum_{y_k \rightarrow y'_k \in \mathcal{R}} \lambda^N_{L,k}(z_\ell) \leq \kappa M^{c_\ell},
\]

where the constant \( c_1 \) is such that \( \| z_\ell - q_L(y'_k - y_k) \|_\infty \leq c_1 M^{c_\ell} \) for any reaction \( y'_k \rightarrow y_k \).

2. By the definition of the projection \( q_\ell \) and \( \lambda^N_{L,k} \)'s, we have

\[
\sum_k \lambda^N_{L,k}(z_\ell - q_L(y'_k - y_k))p^N_{f_{sp}}(z_\ell - q_L(y'_k - y_k), S_{H,M}, t) = \sum_u \lambda_{L,u}(z_\ell - \bar{y}_u u + \bar{y}_u, t)
\]

and

\[
\sum_k \lambda^N_{L,k}(z_\ell)p^N_{f_{sp}}(z_\ell - q_L(y'_k - y_k), S_{H,M}, t) = \sum_u \lambda_{L,u}(z_\ell)p_{f_{sp}}(z_\ell - \bar{y}_u u + \bar{y}_u, t).
\]

For (23), we add 1 and \(-1\) to terms \( \lambda^N_{H,k} \)'s, and we summing over \( z_h \in S_{H,M} \). Then (25), (26) and Lemma 3.1 imply that

\[
\frac{d}{dt}p^N_{f_{sp}}(z_\ell, S_{H,M}, t) \leq \sum_u \lambda_{L,u}(z_\ell - \bar{y}_u u + \bar{y}_u)p^N_{f_{sp}}(z_\ell - q_L(y'_k - y_k), S_{H,M}, t)
\]

\[
- \sum_u \lambda_{L,u}(z_\ell)p^N_{f_{sp}}(z_\ell, S_{H,M}, t) + h(N),
\]

where \( h(N) = c_2\kappa M^{c_\ell+1}/N \) with some \( c_2 > 0 \). Similarly we have

\[
\frac{d}{dt}p^N_{f_{sp}}(z_\ell, S_{H,M}, t) \geq \sum_u \lambda_{L,u}(z_\ell - \bar{y}_u u + \bar{y}_u)p^N_{f_{sp}}(z_\ell - q_L(y'_k - y_k), S_{H,M}, t)
\]

\[
- \sum_u \lambda_{L,u}(z_\ell)p^N_{f_{sp}}(z_\ell, S_{H,M}, t) - h(N),
\]

Since \( S_{L,M} \) is a finite subset of \( \mathbb{Z}_{\geq 0}^{d_s} \), we regard \( p^N_{f_{sp}}(\cdot, S_{H,M}, t) \) and \( p_{f_{sp}}(\cdot, t) \) are vectors in \( \mathbb{R}^{\mid S_{L,M} \mid} \). Then \( p_{f_{sp}} \) is a solution of the following system of differential equations,

\[
\frac{d}{dt}p_{f_{sp}}(t) = A_M p_{f_{sp}}(t),
\]

where \( A_M \) is a matrix for which (29) is equivalent to the system of differential equations (24). Let \( v^N(t) = p^N_{f_{sp}}(t) - p_{f_{sp}} \) and let \( b(N) \) be a \( \mid S_{L,M} \mid \)-dimensional vector each of whose component is \( h(N) \). Then by (27) and (28), we have

\[
L_M v^N(t) - b(N) \leq \frac{d}{dt}v^N(t) \leq L_M v^N(t) + b(N)
\]

(30)
where the inequality is component-wise. Hence by Lemma 3.4 and the upper bound of \( \frac{d}{dt} v(t) \) in (30), we have for each \( t \geq 0 \),

\[
v(t) \leq b(N)(t + t^2|S_{L,M}|A_{\text{max}}),
\]

where \( A_{\text{max}} \) is the largest entry of \( A_M \). Applying Lemma 3.4 with \(-v(t)\), the lower bound of \( \frac{d}{dt} v(t) \) in (30) implies that

\[
v(t) \geq -b(N)(t + t^2|S_{L,M}|A_{\text{max}}).
\]

Finally, an upper bound of \( A_{\text{max}} \leq c_1 M^{e_{\ell}} \) is obtained with (25). Thus we conclude that there exists a constant \( c > 0 \) such that for any subset \( A \subset \mathbb{Z}_{\geq 0}^d \), we have

\[
|p_f^{N}(A, S_{L,M}, t) - p_f^{N}(A, t)| \leq \sum_{z_{\ell} \in A} |v^N(z_{\ell}, t)| \leq |S_{L,M}|h(N)(t + t^2|S_{L,M}|c_2 M^{e_{\ell}}) \\
\leq 2h(N)c_1|S_{L,M}|^2t^2 \leq c \frac{M^{e_\ell+1+2d}}{N}.
\]

\[
\square
\]

Lemma 3.6. Suppose the same conditions in Lemma 3.5. We further assume (CD4). If \( M = f(N) \) such that \( \lim_{N \to \infty} \frac{f(N)^2}{N} = 0 \), then for \( N \) sufficiently large, there exists a constant \( c > 0 \) such that

\[
\sum_{z_{\ell} \in A \cap S_{L,M}} |p^N(z_{\ell}, S_{H,M}, t) - p_f^N(z_{\ell}, S_{H,M}, t)| \leq cK \frac{2^{m+1}}{M^{m-e_{\ell}-1}}, \quad \text{and}
\]

\[
\sum_{z_{\ell} \in A \cap S_{L,M}} |p_f^N(z_{\ell}, t) - p(z_{\ell}, t)| = c \frac{t}{M^{m-e_{\ell}}}
\]

3.2 Main theorem and its application

Theorem 3.7. Suppose for the stochastic processes \( Z^N(t) \) associated with \((S, C, R, \xi^N)\), conditions (CD1), (CD2) and (CD3) hold. Suppose further that the stochastic process associated with \((S_L, C_L, R_L, \xi_L)\) admits a stationary distribution \( \pi \) such that (CD4) holds. Then for each \( t \), the projected process \( q_L(Z^N(t)) \) converges to \( Z(t) \) in distribution. Moreover for each \( t \) and for any \( A \subset \mathbb{Z}_{\geq 0}^d \) and for \( \nu \in (0,1) \), there exists \( C > 0 \) such that

\[
|p^N(A, \mathbb{R}_{\geq 0}^r, t) - p(A, t)| \leq \frac{C}{N^{\nu}} \tag{31}
\]

Proof. Note that for any \( M > 0 \), we have

\[
A \times \mathbb{R}_{\geq 0}^r = ((A \times \mathbb{R}_{\geq 0}^r) \cap S_M) \cup ((A \times \mathbb{R}_{\geq 0}^r) \cap S_M^c) \\
\subset (A \cap S_{L,M} \times S_{H,M}) \cup S_M^c.
\]

Thus we have

\[
|p^N(A, \mathbb{Z}_{\geq 0}^r, t) - p(A, t)| \leq |p^N(A \cap S_{L,M}, S_{H,M}, t) - p(A \cap S_{L,M}, t)| \\
+ p(S_M^c, t) + p(S_{L,M}^c, t).
\]

(32)
Lemma 3.3 in Appendix A shows that
\[ p^N(S^c_M, t) \leq \frac{c_0 t^m}{M^m} \quad \text{and} \quad p^N(S^c_{\ell,M}, t) \leq \frac{c_0 t^m}{M^m}, \] (33)
for any positive integer \( m \) and for some constant \( c_0 > 0 \).

Now we show that the term \( |p^N(A \cap S_{L,M}, S_{H,M}, t) - p(A \cap S_{L,M}, t)| \) in (32) is less than \( \frac{c_1 t^{m+1}}{M^m} \) for any positive integer \( m \) and for some positive constant \( c_1 \). To show this, we break up this term into three parts.

\[
|p^N(A \cap S_{L,M}, S_{H,M}, t) - p(A \cap S_{L,M}, t)| \leq \sum_{z_{\ell} \in A \cap S_{L,M}} |p^N(z_{\ell}, S_{H,M}, t) - p^N_{fsp}(z_{\ell}, S_{H,M}, t)| \\
+ \sum_{z_{\ell} \in A \cap S_{L,M}} |p^N_{fsp}(z_{\ell}, t) - p(z_{\ell}, t)| \\
+ \sum_{z_{\ell} \in A \cap S_{L,M}} |p^N_{fsp}(z_{\ell}, S_{H,M}, t) - p_{fsp}(z_{\ell}, t)|,
\] (34)

where \( p^N_{fsp} \) and \( p_{fsp} \) are the distribution of the FSP models of \( Z^N \) and \( Z \) on \( (S_{L,M}, S_{H,M}) \) and \( S_{L,M} \), respectively. Lemma 3.6 in Appendix B shows that both the first sum and the second sum in (34) are less than \( \frac{c_2 t^m}{M^m} \) for any positive integer \( m \) and for some positive constant \( c_2 \). It is also shown by Lemma (3.5) in B that the third term of (34) is less than \( \frac{c_3 t^{m+1}}{N^r} \).

We choose \( M = N^p \) for sufficiently small \( \rho > 0 \) and choose large enough positive integer \( m \) to conclude the result.

\[ \square \]

**Appendix A : Proof of Lemmas in Section 3.1**

**Proof of Lemma 3.2.** Let \( N_{x,k} \) be independent Poisson processes with intensity \( \lambda_{L,u}(x) \). In [1], it is shown that \( J(t) \) has the following representation.

\[
J(t) = \sum_x \sum_{y_u \rightarrow y_u'} \int_0^t \mathbb{1}_{(Z(s-) = x)} N_{x,k}(ds).
\]

Note that the process \( Z \) can only reach \( J(t) \) different points by time \( t \). Moreover [1, Theorem 2] and (CD4) imply that we have \( E(J(t)) < \infty \) Hence, the summation \( \sum_x \) is a finite summation with \( J(t) \) terms almost surely.

Therefore by applying Jesen's inequality twice, we have

\[
J(t)^{m+1} \leq J(t) \sum_x \left( \sum_{y_u \rightarrow y_u' \in R_L} \int_0^t \mathbb{1}_{(X(s-) = x)} N_{x,k}(ds) \right)^{m+1} \\
\leq J(t)|R| \sum_x \sum_{y_u \rightarrow y_u' \in R_L} \left( \int_0^t \mathbb{1}_{(X(s-) = x)} N_{x,k}(ds) \right)^{m+1}.
\]
Since \( \int_0^t N_{x,k}(ds) = \lambda_k(x)t \), Jensen’s inequality implies that
\[
J(t)^{m+1} \leq t^m J(t)|\mathcal{R}| \sum_x \sum_{\vec{y}_u \rightarrow \vec{y}_u' \in \mathcal{R}_L} \lambda_{L,u}(x)^m \int_0^t \mathbbm{1}_{(X(s-)=x)} N_{x,k}(ds).
\]
Therefore
\[
E(J(t)^m) \leq t^{m+1} \sum_x \sum_{\vec{y}_u \rightarrow \vec{y}_u' \in \mathcal{R}_L} \lambda_{L,u}(x)^{m+1} P(X(t) = x).
\]
By the monotonicity of the chemical master equation, there exists a constant \( c \) such that \( P(X(t) = x) \leq c\pi(x) \) for all \( t \) when \( X(0) \) is a point almost surely. Thus finally we have
\[
E(J(t)^m) \leq c t^{m+1} \sum_x \sum_{\vec{y}_u \rightarrow \vec{y}_u' \in \mathcal{R}_L} \lambda_{L,u}(x)^{m+1} \lambda^{m+1}(x)\pi(x).
\]
Since \( \lambda_{L,u}(x) \) is a polynomial for each \( \vec{y}_u \rightarrow \vec{y}_u' \in \mathcal{R}_L \), the proof completes by (CD4).

**Proof of Lemma 3.3.** Let \( Path_M \) be a collection of all possible \( M \) consecutive reactions of \( \mathcal{R} \) for \( Z^N \) started at \( Z^N(0) \). Each element \( \eta \in Path_M \) is an ordered pair of \( M \) reactions in \( \mathcal{R} \). That is, \( \eta = \{y(\eta, 1) \rightarrow y'(\eta, 1), \ldots, y(\eta, M) \rightarrow y'(\eta, M)\} \), where \( y(\eta, i) \rightarrow y'(\eta, i) \in \mathcal{R} \) for each \( i \). We define \( w(\eta, j) \) be a point after \( j \) consecutive jumps in \( \eta \) from \( Z^N(0) = (x_\ell, x_h) \). That is
\[
w(\eta, j) = (x_\ell, x_h) + \sum_{i=1}^j (y'('i, i) - y('i, i)) = (w(\eta, j)_\ell, w(\eta, j)_h)
\]
for \( j = 1, 2, \ldots, M \), where
\[
w(\eta, j)_\ell = \left(x_\ell + \sum_{i=1}^j q_L(y'('i, i) - y('i, i))\right) \quad \text{and} \quad w(\eta, j)_h = \left(x_h + \sum_{i=1}^j q_H(y'_a('i, i) - y_a('i, i))\right).
\]
Let \( A_\eta \) represent the event of \( M \) consecutive jumps for \( Z^N \) along the reactions in \( \eta \in Path_M \) in its order. Let also \( \hat{A}_\eta \) represent the event of \( M \) consecutive jumps for \( Z \) along the ordered reactions \( q_L(y(\eta, 1)) \rightarrow q_L(y'(\eta, 1)), \ldots, q_L(y(\eta, M)) \rightarrow q_L(y'(\eta, M)) \).

In the following main steps of this proof, we take three key ideas into account. First, under the event \( A_\eta \), the stopping time \( \tau_M \) is exponentially distributed with rate \( \lambda(w(\eta, j)) \) [20], where
\[
\lambda(w(\eta, j)) = \sum_{y_k \rightarrow y'_k \in \mathcal{R}} \lambda^N_k(w(\eta, i)).
\]
More precisely,
\[
P^N(\tau^N_M < t | A_\eta) = P \left( \sum_{i=1}^M T^N(i) < t \right),
\]
where \( T^N \)'s are independent exponential distributions with rate \( \lambda^N(w(\eta, i)) \).

Secondly, by Lemma 3.1, we have for sufficiently large \( N \),
\[
\lambda^N(z_\ell, z_h) \leq 2\lambda_L(z_\ell) \quad \text{for each} \ (z_\ell, z_h) \in (S_{L,M}, S_{H,M}),
\]
where $\lambda_L(z) = \sum_{g_u \rightarrow g_u' \in R_L} \lambda_{L,u}(z)$. This combined with (36), we have

$$P_N^N(T_M^N < t | A_\eta) \leq P \left( \sum_{i=1}^{M} T(i) < t/2 \right) = P(\tau_M < t/2 | \tilde{A}_\eta),$$

where $T_i$'s are independent exponential distributions with rate $\lambda_L(w(\eta, i)e)$.

Finally note that for each $\eta \in Path_M$, the event $A_\eta$ occurs if and only if reaction $y(\eta, i) \rightarrow y'(\eta, i)$ fires at each state $w(\eta, i)$ out of all reactions in $R$. This implies that

$$P(A_\eta) = \prod_{i=1}^{M} \frac{\lambda^N_{\eta,i}(w(\eta, i))}{\lambda^N(w(\eta, i))},$$

where $\lambda^N_{\eta,i}$ is the intensity function of the reaction $y(\eta, i) \rightarrow y'(\eta, i)$. As shown in Lemma 3.1, we split $\lambda^N_{\eta,i}(w(\eta, i))$ into two parts as

$$\lambda^N_{\eta,i}(w(\eta, i)) = \lambda^N_{L,\eta,i}(w(\eta, i)e)\lambda^N_{H,\eta,i}(w(\eta, i)h),$$

where $\lambda^N_{L,\eta,i}(z_\ell) = \kappa_\ell z_\ell^{q_d(\eta(\eta, i))}$ and $\lambda^N_{H,\eta,i}(z_h) = z_h^{q_d(\eta(\eta, i))}$ for each $z = (z_\ell, z_h)$. Therefore Lemma 3.1 and (38) imply that

$$P_N^N(A_\eta) \leq \left( \frac{1 + cM}{1 - cM} \right)^M \prod_{i=1}^{M} \frac{\lambda_{\eta,i}(w(\eta, i)e)}{\lambda_L(w(\eta, i)e)} = \left( \frac{1 + cM}{1 - cM} \right)^M P(\tilde{A}_\eta),$$

for some $c > 0$. Hence by (37) and (39), we have

$$P_N^N(\tau_M^N < t) = \sum_{\eta \in Path_M} P_N^N(\tau_M^N < t | A_\eta) P_N^N(A_\eta)$$

$$\leq \left( \frac{1 + cM}{1 - cM} \right)^M \sum_{\eta \in Path_M} P(\tau_M < t/2 | \tilde{A}_\eta) P(\tilde{A}_\eta)$$

$$= \left( \frac{1 + cM}{1 - cM} \right)^M P(\tau_M < t/2).$$

To show the result (20), let $J(t) = \sup\{n : \tau_n \leq t\}$ be the number of jumps of $Z$ by time $t$. Then, applying the Chebyshev’s inequality, we have

$$P(\tau_M < t/2) \leq P(J(t/2) > M) \leq \frac{E(J(t/2)^m)}{M^m}.$$

Then (20) follows from Lemma 3.2.

The result (21) follows from the fact that

$$\lim_{N \to \infty} \left( 1 + \frac{M}{N} \right)^M = 1 \quad \text{and} \quad \lim_{N \to \infty} \left( 1 - \frac{M}{N} \right)^M = 1,$$

if $\lim_{N \to \infty} \frac{M^2}{N} = \lim_{N \to \infty} \frac{f(N)^2}{N} = 0$.  

\[\square\]
Proof of Lemma 3.4. $\frac{d}{dt} v(t) \leq L v(t) + b$ can be written as

$$v(t) \leq tb + \int_0^t L v(s) ds$$

allowing the inequality holds component-wisely. Then by the multivariable Gronwall’s inequality [8],

$$v(t) \leq tb + t \int_s^t V(t, s) L b ds,$$

where $V(t, s)$ satisfies

$$V(t, s) = I + \int_s^t LV(x, s) dx,$$

allowing the equality holds component-wisely. By taking time-derivative, we notice that $i$ th column of $V(t, s)$ is a solution $u$ of the system of differential equation (22) with $u_i(0) = 1$ and $u_j(0) = 0$ otherwise. Therefore by the assumption of $L$, every component of $V(t, s)$ is positive and less than or equal to 1 for $t \geq s$, so it implies that $V(t, s)Lb \leq nL_{\text{max}}b$ for $t \geq s$. Thus for each $t$, the result follows.

Proof of Lemma 3.6. We define two subsets of the state space of $Z^N$ and $Z$, respectively.

$$G_M = \{(z_t, z_h) \in (S_{L,M}, S_{H,M}) : (z_t, z_h) + y_k^N - y_k^N \not\in (S_{L,M}, S_{H,M}) \text{ for some } y_k \to y'^k \in \mathcal{R}\},$$

$$G_{L,M} = \{z_t \in S_{L,M} : z_t + \tilde{y} - \tilde{y} \not\in S_{L,M} \text{ for some } \tilde{y} \to \tilde{y}' \in \mathcal{R}_L\}.$$

By [19], we have

$$\sum_{z_t \in A \cap S_{L,M}} |p^N(z_t, S_{H,M}, t) - p^N_{fsp}(z_t, S_{H,M}, t)| \leq \int_0^t \sum_{(z_t, z_h) \in G_M} \sum_k \lambda^N_{H,k}(z_h) \lambda^N_{L,k}(z_t) p^N(z_t, z_h, s) ds$$

For each reaction $y_k \to y'_k$ and $(z_t, z_h) \in G_M$.

the intensities $\lambda^N_{H,k}(z_h)$ and $\lambda^N_{L,k}(z_t)$ have upper bounds by Lemma 3.1 and (25) as

$$\lambda^N_{H,k}(z_h) \leq 2 \quad \text{and} \quad \sum_k \lambda^N_{L,k}(z_t) \leq c_1 \kappa M^{c_1},$$

with some constant $c_1 > 0$. Moreover, for $M$ large enough, $Z^N(s) \in G_M$ implies that there are at least $M/2$ jumps of $Z^N$ by time $s$. Therefore Lemma 3.3 replacing $M$ by $M/2$ implies that for any $m > 0$, there exists a constant $c_2 > 0$ such that

$$P(Z^N(s) \in G_M) \leq P(\tau_{M/2}^N < s) \leq \frac{c_2 2^m s^m}{M^m}.$$ 

Noting that $|G_M| \leq c_3 (d + r) M$ for some $c_3 > 0$, we have that for $N, M$ sufficiently large, there exists a constant $c' > 0$ such that

$$\sum_{z_t \in A \cap S_{L,M}} |p^N(z_t, S_{H,M}, t) - p^N_{fsp}(z_t, S_{H,M}, t)| \leq c' M|G_M| \frac{2^m m!^m + 1}{M^{m-c_t}}$$

$$\leq c' \frac{2^m m!^m + 1}{M^{m-c_t-1}}.$$
For the term \( \sum_{z \in A \cap S_{L,M}} |p_{fsp}(z, t) - p(z, t)| \), we use the monotonicity argument of system of ordinary differential equations. Let \( c_4 > 0 \) be a constant such that \( p(z, 0) \leq c_4 \pi(z) \) for all \( z \in \mathbb{Z}_d^d \). Then by the monotonicity argument for the Kolmogorov equation, we have \( p(z, t) \leq \gamma \pi(z) \) for any \( z \) and for any \( t \). Theorem in [19] combined with the monotonicity of \( p(z, t) \), there exists a constant \( c'' > 0 \) such that

\[
\sum_{z \in A \cap S_{L,M}} |p_{fsp}(z, t) - p(z, t)| \leq t c_4 \sum_{z \in G_{L,M}} \sum_{u} \lambda_{L,u}(z) p(z, s) ds
\]

\[
\leq t c_4 \sum_{z \in G_{L,M}} \sum_{u} \lambda_{L,u}(z) \pi(z)
\]

\[
\leq t c'' M^{\ell} \max_{z \in G_{L,M}} \pi(z).
\]

By ((CD4)), for any \( m > 0 \), we have \( \max_{z \in G_{L,M}} \pi(z) \leq \frac{1}{M^m} \) for \( M \) sufficiently large. Thus with \( c = \max\{c', c''\} \), the proof completes.

**Appendix B : Relation between two different system size scales**

In this section we show that when the reactant of each reaction contains at least one species of high copy number, the system scaling regime used in this paper and the system scaling regime used in [3] are equivalent under certain time-scaling.

Note that we denote \( S_1, \ldots, S_d \) for species of low initial copy number (constant order) and denote \( S_{d+1}, \ldots, S_{d+r} \) for species of high initial copy number (order \( N \)). In [3], the rate constant \( \bar{\kappa}_{y \rightarrow y'}^{N} \) for each reaction \( y \rightarrow y' \) is scaled as

\[
\bar{\kappa}_{y \rightarrow y'}^{N} = \begin{cases} \frac{\kappa_{y \rightarrow y'}}{N^{1+1/\|u(y)\|_1}} & \text{if } \|u(y)\|_1 > 0, \\ \kappa_{y \rightarrow y'} & \text{otherwise}, \end{cases} \tag{40}
\]

for some constant \( \kappa_{y \rightarrow y'} \). For a reaction network \((S, \mathcal{C}, \mathcal{R})\), we denote \( \bar{\kappa}^N \) the set of rate constants under the scaling in (40). Then we denote \( \bar{\lambda}_{y \rightarrow y'}^{N} \) the intensity of a reaction \( y \rightarrow y' \) for the network system \((S, \mathcal{C}, \mathcal{R}, \bar{\kappa}^N)\). Note that

\[
\bar{\lambda}_{y \rightarrow y'}^{N}(x) = \frac{1}{N} \lambda_{y \rightarrow y'}^{N}(x) \tag{41}
\]

if \( \|u(y)\|_1 > 0 \) by the definitions.

**Theorem 3.8.** Let \((S, \mathcal{C}, \mathcal{R})\) be a reaction network with \( S = \{S_1, \ldots, S_d, S_{d+1}, \ldots, S_{d+r}\} \). Let \( X^N \) and \( \bar{X}^N \) be Markov processes associated to network systems \((S, \mathcal{C}, \mathcal{R}, \bar{\kappa}^N)\) and \((S, \mathcal{C}, \mathcal{R}, \kappa^N)\), respectively. Suppose \( \|u(y)\|_1 > 0 \) for the source complex \( y \) in each reaction \( y \rightarrow y' \). Then if \( X^N(0) = \bar{X}^N(0) \) for each \( N \), then for each \( t \geq 0 \), \( X^N(t) \) is equal to \( \bar{X}^N(t/N) \) in distribution.

**Proof.** By the random time change represetiona [11],

\[
\bar{X}^N(t) = \bar{X}^N(0) + \sum_{y \rightarrow y' \in \mathcal{R}} \bar{Y}_{y \rightarrow y'} \left( \int_0^t \bar{\lambda}_{y \rightarrow y'}(\bar{X}^N(s)) ds \right) (y' - y), \tag{42}
\]

\[
Y_{y \rightarrow y'}(\bar{X}^N(t)) = \int_0^t \lambda_{y \rightarrow y'}(X^N(s)) ds, \tag{43}
\]
where $\bar{Y}_{y\to y'}$’s are independent unit Poisson random variables.

Since $\|u(y)\|_1 > 0$ for all source complexes, (41) implies that

$$X^N(t/N) = X^N(0) + \sum_{y\to y' \in \mathcal{R}} \bar{Y}_{y\to y'} \left( \int_0^t \bar{\lambda}^N_{y\to y'}(X^N(s/N))ds \right) (y' - y)$$

with change the time variable $t$ by $t/N$ in (42). This is equal in distribution to the representation of $X^N$ such that

$$X^N(t) = X^N(0) + \sum_{y\to y' \in \mathcal{R}} Y_{y\to y'} \left( \int_0^t \lambda^N_{y\to y'}(X^N(s))ds \right) (y' - y).$$

where $Y_{y\to y'}$’s are independent unit Poisson random variables.

References

[1] David F Anderson, Daniele Cappelletti, Masanori Koyama, and Thomas G Kurtz. Non-explosivity of stochastically modeled reaction networks that are complex balanced. *Bulletin of mathematical biology*, 80(10):2561–2579, 2018.

[2] David F Anderson, Daniele Cappelletti, and Thomas G Kurtz. Finite time distributions of stochastically modeled chemical systems with absolute concentration robustness. 2016.

[3] David F Anderson, Daniele Cappelletti, and Thomas G Kurtz. Finite time distributions of stochastically modeled chemical systems with absolute concentration robustness. *SIAM Journal on Applied Dynamical Systems*, 16(3):1309–1339, 2017.

[4] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.*, 72(8):1947–1970, 2010.

[5] Karen Ball, Thomas G Kurtz, Lea Popovic, Greg Rempala, et al. Asymptotic analysis of multiscale approximations to reaction networks. *The Annals of Applied Probability*, 16(4):1925–1961, 2006.

[6] Youfang Cao and Jie Liang. Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. *BMC Systems Biology*, 2(1):30, 2008.

[7] Youfang Cao, Anna Terebus, and Jie Liang. Accurate chemical master equation solution using multi-finite buffers. *Multiscale Modeling & Simulation*, 14(2):923–963, 2016.

[8] Jagdish Chandra and Paul W. Davis. LINEAR GENERALIZATIONS OF GRONWALL’S INEQUALITY. *Proceedings of the American Mathematical Society*, 60, 1976.

[9] German Enciso and Jinsu Kim. Embracing noise in chemical reaction networks. *Bulletin of mathematical biology*, 81(5):1261–1267, 2019.
[10] German A. Enciso. Transient absolute robustness in stochastic biochemical networks. *J. R. Soc. Interface*, 13:20160475, 2016.

[11] Stewart N Ethier and Thomas G Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, New York, 1986.

[12] Ankit Gupta and Mustafa Khammash. A finite state projection algorithm for the stationary solution of the chemical master equation. *The Journal of Chemical Physics*, 147(15), 2017.

[13] Hye-Won Kang, Wasiur R. KhudaBuksh, Heinz Koeppl, and Grzegorz A. Rempala. Quasi-steady-state approximations derived from the stochastic model of enzyme kinetics. *Bull. Math. Bio.*, 2019.

[14] Hye-Won Kang and Thomas G Kurtz. Separation of time-scales and model reduction for stochastic reaction networks. *Annals of Applied Probability*, 23(2):529–583, 2013.

[15] Jae Kyoung Kim, Grzegorz A Rempala, and Hye-Won Kang. Reduction for stochastic biochemical reaction networks with multiscale conservations. *Multiscale Modeling & Simulation*, 15(4):1376–1403, 2017.

[16] Jae Kyoung Kim and Eduardo D Sontag. Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. *PLoS computational biology*, 13(6):e1005571, 2017.

[17] Thomas G Kurtz. The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *J. Chem. Phys.*, 57(7):2976–2978, 1972.

[18] Brian Munsky. The finite state projection approach for the solution of the master equation and its applications to stochastic gene regulatory networks. Ph.D. thesis, 2008.

[19] Brian Munsky and Khammash Mustafa. The finite state projection approach for the analysis of stochastic noise in gene networks Dissertation. *IEEE Trans. Autom. Contr*, 53:201–214, 2008.

[20] James Norris. *Markov Chains*. Cambridge University Press, 1997.