A multi-time-scale analysis of chemical reaction networks: II. Stochastic systems

Xingye Kan1 · Chang Hyeong Lee2 · Hans G. Othmer3

Received: 17 August 2015 / Revised: 4 December 2015 / Published online: 5 March 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract We consider stochastic descriptions of chemical reaction networks in which there are both fast and slow reactions, and for which the time scales are widely separated. We develop a computational algorithm that produces the generator of the full chemical master equation for arbitrary systems, and show how to obtain a reduced equation that governs the evolution on the slow time scale. This is done by applying a state space decomposition to the full equation that leads to the reduced dynamics in terms of certain projections and the invariant distributions of the fast system. The rates or propensities of the reduced system are shown to be the rates of the slow reactions conditioned on the expectations of fast steps. We also show that the generator of the reduced system is a Markov generator, and we present an efficient stochastic simulation algorithm for the slow time scale dynamics. We illustrate the numerical accuracy of the approximation by simulating several examples. Graph-theoretic techniques are

All authors contributed equally to this work.

Supported in part by NSF Grants DMS # 9517884 and 131974 and NIH Grant # GM 29123 to H. G. Othmer and by National Research Foundation of Korea (2014R1A1A2054976) to C. H. Lee.

Xingye Kan
xkan@umn.edu
Chang Hyeong Lee
chlee@unist.ac.kr
Hans G. Othmer
othmer@math.umn.edu

1 School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA
2 Ulsan National Institute of Science and Technology, Ulsan Metropolitan City 698-798, South Korea
3 School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA
used throughout to describe the structure of the reaction network and the state-space transitions accessible under the dynamics.

**Keywords**  Stochastic dynamics · Reaction networks · Graph theory · Singular perturbation

**Mathematics Subject Classification** 92C45 · 80A30 · 34E13

1 Introduction and background

While singular perturbations techniques and the quasi-steady-state approximation (QSSA) have a long history of use in deterministic descriptions of chemical reaction kinetics (cf. Lee and Othmer 2009 for a review), it was apparently not applied to discrete stochastic descriptions of chemical kinetics until Janssen proposed a method for adiabatic elimination of fast variables in stochastic chemical reaction networks (Janssen 1989a, b). Janssen began with a master equation description and, using projection techniques related to volume expansions, obtained a reduced master equation and showed in some examples that an intermediate chemical species can be eliminated in a network if a reaction occurs much faster than the others. Moreover, he also gave examples in which the master equation cannot be reduced via these techniques (Janssen 1989b). It is sometimes assumed that the results of a deterministic reduction of a reaction network yields correct results for the stochastic description, but examples show that this is not correct (Thomas et al. 2011).

The Gillespie algorithm (Gillespie 2007) is the most widely-used algorithm for simulating stochastic reactions but it can be very inefficient when there are multiple time scales in the reaction dynamics. Stochastic simulation algorithms of multiscale reaction networks were developed more or less simultaneously by Haseltine and Rawlings (2002) and Rao and Arkin (2003). The former authors formulated the changes in the species numbers \( n \) in terms of extents \( \eta \) by defining \( n(t) = n(0) + \nu \mathcal{E} \eta(t) \), and divided the reaction extents into those of slow reactions and fast reactions. Rao and Arkin (2003) divided the set of species into ‘primary’ \( (y) \) and ‘intermediate or ephemeral’ \( (z) \), and assumed two conditions on the conditional variable \( z|y \), (i) the conditional variable \( z|y \) is Markovian and (ii) \( z|y \) is at quasi-steady-state on the slow time scale. However, the first assumption was not justified and one cannot classify species as slow and fast—rather it is reactions that are slow or fast and species can participate in both. Mastny et al. (2007) applied singular perturbation analysis to remove the QSS species for a number of model systems having small populations, and for networks where non-QSS species have large populations, they also utilized the \( \Omega \)-expansion to reduce the master equation. More recent work deals with the role played by the reduction on the level of noise in the solution (Srivastava et al. 2011).

Several others have proposed and implemented variations of these two approaches for hybrid simulations of stochastic systems. These include partitioning the system dynamically (Salis and Kaznessis 2005) and solving the fast reactions using an SDE.
approximation. The effective reaction rate expressions for Hill-type kinetics derived from singular perturbation analysis of the deterministic system of equations have been used in stochastic simulations and the results compared to simulations for the full system (Bundschuh et al. 2003). There is also a partitioning method based on the variance of species, which led to a hybrid method coupling deterministic (small variance) and stochastic (large variance) by Hellander and Lötstedt (2007). Goutsias showed that fast reactions can be eliminated when probabilities of slow reactions depend at mostly linearly on the states of fast reactions (Goutsias 2005). Utilizing the degree of advancement or extent, he followed Haseltine and Rawling’s approach in order to separate variables into fast and slow variables. By utilizing a Taylor series expansion he showed that when the slow transition rates or propensity functions depend linearly on fast extents, the fast reaction kinetics of a stochastic biochemical system can be approximated by the conditional mean of extents for the fast kinetics. Although his approach is rigorous, it has some drawbacks. In many nonlinear systems, slow reactions depend nonlinearly on fast reactions. For example, if a slow reaction subsystem consists of bimolecular reactions with two different reactants which are affected by fast reactions, his method cannot be applied due to the nonlinear dependence of the slow reactions on fast reactions. Peles et al. (2006) applied singular perturbation theory and finite state projection method to obtain an approximate master equation for certain linear reaction systems, where clusters of fast-transitioned states have been identified. A rigorous mathematical framework is lacking in their derivation, which we provide herein.

Computational methods have also been developed by many groups. Cao et al. (2005) proposed a slow scale stochastic simulation algorithm (SSA). They first identified fast and slow reaction channels and defined fast and slow species according as the species are changed by fast reactions or not. Assuming that the fast processes are stable, they approximated the propensity functions on a slow time scale, and showed that computation of the effective propensity functions requires first and second moments of the steady-state distribution of the partitioned fast reaction subsystem. They illustrated numerical results for a system with one independent fast species. Using a partitioning approach similar to that in Haseltine and Rawlings (2002), Rao and Arkin (2003) Cao et al. (2005), Chevalier and EI-Samad (2009) derived evolution equations for both fast and slow reactions that led to an SSA with slow reaction trajectories in which the SSA is run until the first occurrence of a slow reaction. E et al. (2005) proposed a nested stochastic simulation algorithm with inner loops for the fast reactions and outer loops for the slow reactions. Strong convergence of the nested SSA has been proved recently by Huang and Liu (2014) and a speed up of the algorithm is proposed by using the tau-leaping method as the inner solver. More recently, Kim et al (2014) investigated the validity of the stochastic QSSA, where the propensity functions resulted from their deterministic counterpart. Under the moment closure assumption, they have shown that the stochastic QSSA is accurate for a two-dimensional monostable system, if the deterministic QSS solution is not very sensitive. Goutsias and Jenkinson (2013) provide an excellent review article covering recently developed techniques on approximating and solving master equations for complex networks. Other recent papers include (Cotter 2015; Smith et al. 2015).

As described above, many computational results and some algorithms have been reported to date, but a rigorous analysis of stochastic reaction networks that involve
multiple time scales based on singular perturbation has heretofore not been done. Preliminary work on this by one of the authors (Lee and Lui 2009) proposed a reduction method based on a singular perturbation analysis for networks with two or more time scales under the assumption that the sub-graph of fast steps in the network is strongly connected. Our objectives here are to develop a rigorous analytic framework for the reduction of general stochastic networks with two widely-separated time scales, to prove that the generator of the reduced system is Markovian, and to illustrate the numerical accuracy and speedup of the reduction method for several biological models.

2 The general formulation for reacting systems

2.1 Deterministic evolution equations

To set notation for the stochastic analysis, we first recall some notation used in our previous analysis of deterministic systems (Lee and Othmer 2009), hereafter referred to as I. Further details can be found there.

Suppose that the reacting mixture contains the set $\mathcal{M}$ of $m$ chemical species $\mathcal{M}_i$ that participate in a total of $r$ reactions. Let $v_{i,\ell}$ be the stoichiometric coefficient of the $i$th species in the $\ell$th reaction. The $v_{i,\ell}$ are non-negative integers that represent the normalized molar proportions or stoichiometric coefficients of the species in a reaction. Each reaction is written in the form

$$
\sum_i v_{i,\ell}^{\text{reac}} \mathcal{M}_i \rightarrow \sum_i v_{i,\ell}^{\text{prod}} \mathcal{M}_i, \quad \ell = 1, \ldots, r,
$$

where the sums are over reactants and products, respectively in the $\ell$th reaction. In this formulation, the forward and reverse reaction of a reversible pair are considered separately, as two irreversible reactions. Once the reactants and products for each reaction are specified, the significant entities so far as the network topology is concerned are not the species themselves, but rather the linear combinations of species that appear as reactants or products in the various elementary steps. These linear combinations of species are complexes (Horn and Jackson 1972), and we suppose that there are $p$ of them. A species may also be a complex as is the case for first-order reactions. Once the complexes are fixed, their composition is specified unambiguously, and we let $v$ denote the $m \times p$ matrix whose $j$th column encodes the stoichiometric amounts of the species in the $j$th complex.

The set of reactions gives rise to a directed graph $\mathcal{G}$ as follows. Each complex is identified with a vertex $V_j$ in $\mathcal{G}$ and a directed edge $E_\ell$ is introduced into $\mathcal{G}$ for each reaction. The topology of $\mathcal{G}$ is encoded in its vertex-edge incidence matrix $\mathcal{E}$, which is defined as follows.

$$
\mathcal{E}_{j,\ell} = \begin{cases} 
+1 & \text{if } E_\ell \text{ is incident at } V_j \text{ and is directed toward it} \\
-1 & \text{if } E_\ell \text{ is incident at } V_j \text{ and is directed away from it} \\
0 & \text{otherwise}
\end{cases}
$$

(2.2)
Since there are $p$ complexes and $r$ reactions, $E$ has $p$ rows and $r$ columns, and every column has exactly one $+1$ and one $-1$. Each edge carries a nonnegative weight $R_{\ell}(c)$ given by the intrinsic rate of the corresponding reaction. For example, the following table gives four classes of first-order reactions studied in Gadgil et al. (2005) and two additional bimolecular reaction types.

For either type III or VI reactions there may also be different types of products, e.g., $A \rightarrow B + C$ may represent the decomposition of a complex. Inclusion of such types poses no difficulties, but if such reactions are reversible we restrict the type to uni- or bimolecular reactions.

If at least one reaction of type I, II, IV, or V is present the stoichiometric matrix is

$$\nu = \begin{bmatrix} I & 0 \end{bmatrix}. \tag{2.3}$$

wherein the column of zeroes in the complex matrix represents the null complex $\phi$, which by definition contains no time-varying species. If the system is closed and the reactions are all first order

$$\nu = \begin{bmatrix} I \end{bmatrix}. \tag{2.4}$$

If all species are also complexes, which can occur when there are inputs ($\emptyset \rightarrow M_i$), outputs ($M_i \rightarrow \emptyset$), and first-order decay or conversion reactions ($M_j \rightarrow M_i$), the stoichiometric matrix has the form

$$\nu = \begin{bmatrix} I & \nu_1 & 0 \end{bmatrix}. \tag{2.5}$$

wherein $\nu_1$ defines the stoichiometry of the higher-order complexes. In this case the corresponding incidence matrix $E$ can be written as follows.

$$E = \begin{bmatrix} E_1 & E_2 & E_o \end{bmatrix} \tag{2.6}$$

where $E_1$ represents first-order reactions, $E_2$ represents second-order reactions, and $E_o$ represents input and output steps, all of the appropriate dimensions. An alternate form of the complex and incidence matrices arises if the inputs or outputs are only of type I or II, for then the null complex $\phi$ can be omitted from $\nu$, the $\pm 1$’s omitted from $E$, and the inputs or outputs represented by a separate vector in the evolution equations given below. In either case the stoichiometry of the reactions and the topology of the network are easily encoded in $\nu$ and $E$, respectively.

In this notation the evolution of the composition of a reacting mixture is governed by

$$\frac{dc}{dt} = \nu E R(c), \quad c(0) = c_0 \tag{2.6}$$

where the $jth$ column of $\nu$ gives the composition of the $jth$ complex and $R_{\ell}(c)$ is the rate of the $\ellth$ reaction, or equivalently, the flow on the $\ellth$ edge of $G$. The matrix $\hat{\nu} \equiv \nu E$ is called the stoichiometric matrix when the composition of complexes and the topology of $G$ are not encoded separately, as we do here (Aris 1965). One can interpret the factored form in (2.6) as follows: the vector $R$ gives the flows on edges
due to reactions of the complexes, the incidence matrix maps this flow to the sum of all flows entering and leaving a given node (a complex), and the matrix $v$ converts the net change in a complex to the appropriate change in the molecular species.

A **component** is a connected subgraph $G_1 \subset G^0$ that is maximal with respect to the inclusion of edges, i.e., if $G_2$ is a connected subgraph and $G_1 \subset G_2 \subset G^0$, then $G_1 = G_2$. An isolated vertex is a component and every vertex is contained in one and only one component. A directed graph $G$ is **strongly connected** if for every pair of vertices $(V_i, V_j)$, $V_i$ is reachable from $V_j$ and vice-versa. Thus a directed graph is strongly connected if and only if there exists a closed, directed edge sequence that contains all the edges in the graph. A **strongly-connected component** of $G$ (a strong component or SCC for short) is a strongly-connected subgraph of a directed graph $G$ that is maximal with respect to inclusion of edges. An isolated vertex in a directed graph is a strong component, and every vertex is contained in one and only one component. Strong components in the directed graph $G$ are classified into three distinct types: sources, internal strong components and absorbing strong components.

A source is a subgraph which has outgoing edges to other strong components and has no incoming edges from other strong components. An internal strong component is a strong component in which edges from other strong components terminate and from which edges to other strong components originate. An absorbing strong component is a strong component from which no edges to other strong components originate. If $G$ has $p$ vertices and $q$ strong components then it is easily shown that the rank of $E$ is $\rho(E) = p - q$ (Chen 1971).

For ideal mass-action kinetics, which we consider here, the flow on the $\ell$th edge, which originates at the $j$th vertex, depends only on the species in the $j$th complex, and the rate can be written as

$$R_\ell(c) = k_{\ell j} P_j(c) \quad \text{where} \quad P_j(c) = \prod_{i=1}^{m} (c_i)^{v_{ij}}$$

for every reaction that involves the $j$th complex as the reactant. Thus the rate vector can be written as

$$R(c) = K P(c)$$

where $K$ is an $r \times p$ matrix with $k_{\ell j} > 0$ if and only if the $\ell$th edge leaves the $j$th vertex, and $k_{\ell j} = 0$ otherwise. Since each row of $K$ has one and only one positive entry, we now denote the only positive entry in $\ell$th row by $k_{\ell}$. The topology of the underlying graph $\hat{G}$ enters into $K$ as follows. Define the exit matrix $E_e^\ell$ of $\hat{G}$ by replacing all 1’s in $E$ by zeroes, and changing the sign of the resulting matrix. Let $\hat{K}$ be the $r \times r$ diagonal matrix with the $k_{\ell}$’s, $\ell = 1, \ldots, r$, along the diagonal. Then it is easy to see that $K = \hat{K} E_e^T$ and therefore

$$\frac{dc}{dt} = v E \hat{K} E_e^T P(c)$$

This form also includes non-ideal mass action rate laws, but the concentrations in (2.7) are then replaced by the activities of the species in the reactant complex, and as a result the flow on a edge may depend on all species in the system.
It follows from the definitions that (i) the \((p,q)\)th entry, \(p \neq q\), of \(\mathcal{E}\hat{K}\mathcal{E}^T\) is nonzero (and positive) if and only if there is a directed edge \((q,p) \in \mathcal{G}\), (ii) each diagonal entry of \(\mathcal{E}\hat{K}\mathcal{E}^T\) is minus the sum of the \(k\)’s for all edges that leave the \(j\)th vertex, and (iii) the columns of \(\mathcal{E}\hat{K}\mathcal{E}^T\) all sum to zero, and so the rank of \(\mathcal{E}\hat{K}\mathcal{E}^T\) is \(\leq p - 1\).

If one separates the inputs, which are constants, one can write this as

\[
\frac{dc}{dt} = v\mathcal{E}\hat{K}\mathcal{E}^T P(c) + \Phi,
\]

where \(\Phi\) is the constant input and both \(v\) and \(\mathcal{E}\) are modified appropriately. Herein we use the evolution equations in the form (2.9) unless stated otherwise.

One can also describe the evolution of a reacting system in terms of the number of molecules present for each species. Let \(n = (n_1, n_2, \ldots, n_m)\) denote the discrete composition vector whose \(i\)th component \(n_i\) is the number of molecules of species \(\mathcal{M}_i\) present in the volume \(V\). This is related to the composition vector \(c\) by \(n = \mathcal{N}_A V c\), where \(\mathcal{N}_A\) is Avagadro’s number, and although the \(n_i\) take discrete values, they are regarded as continuous when large numbers are present. From (2.6) we obtain the deterministic evolution for \(n\) as

\[
\frac{dn}{dt} = v\mathcal{E}\tilde{R}(n)
\]

where \(\tilde{R}(n) \equiv \mathcal{N}_A V R(n/\mathcal{N}_A V)\). In particular, for ideal mass-action kinetics

\[
\tilde{R}_\ell(n) = \mathcal{N}_A V k_\ell P_j(n/\mathcal{N}_A V)
\]

\[
= \mathcal{N}_A V k_\ell \prod_{i=1}^m \left(\frac{n_i}{\mathcal{N}_A V}\right)^{v_{ij}} = \frac{k_\ell}{(\mathcal{N}_A V)^{\sum_i v_{ij}}} \prod_{i=1}^m (n_i)^{v_{ij}} = \hat{k}_\ell \prod_{i=1}^m (n_i)^{v_{ij}}.
\]

2.2 Invariants of reaction networks

The kinematic invariants and the kinetic invariant manifolds in a deterministic description of reactions in a constant-volume system are discussed in detail in (Othmer 1979). In general the concentration space has the decomposition

\[
\mathbf{R}_m = \mathcal{N}[(v\mathcal{E})^T] \oplus \mathcal{R}[v\mathcal{E}],
\]

where \(\mathcal{N}(A)\) denotes the null space of \(A\), and \(\mathcal{R}(A)\) denotes the range of \(A\). The solution of (2.6), which defines a curve in \(\mathbf{R}_m\) through an initial point \(c_0\), can be written

\[
c(t) = c_0 + v\mathcal{E} \int_0^t R(c(\tau))d\tau.
\]
This shows that \( c(t) - c_0 \in \mathcal{R}(vE) \), and the intersection of the translate of \( \mathcal{R}(vE) \) by \( c_0 \), which formally is a coset of \( \mathcal{R}(vE) \) with the non-negative cone \( C^+_m \) of \( \mathbb{R}^m \), defines the reaction simplex \( \Omega(c_0) \). While this terminology has a long history (Aris 1965), \( \Omega \) is a simplex in the mathematical sense only if the intersection of the coset with \( C^+_m \) is compact, which occurs if and only if there is a vector \( y > 0 \in \mathcal{N}[(vE)^T] \) (Othmer 1979). This is only guaranteed in closed systems, where the total mass is conserved and \( y \) comprises the molecular weights of the species, and therefore in general we should call \( \Omega \) the kinetic manifold, but we retain the standard terminology.

First suppose that the system is closed – the case of an open system is discussed later. A vector \( a \in \mathbb{R}^m \) defines an invariant linear combination of concentrations if

\[
\langle a, vE R(c) \rangle = 0,
\]

and these are called kinematic invariants if \( a \in \mathcal{N}[(vE)^T] \) (Othmer 1979). The following important properties of these invariants are proven in (Lee and Othmer 2009).

**P1** One can choose a basis for \( \mathcal{N}[(vE)^T] \) of vectors with integer entries.

**P2** If the reaction simplex \( \Omega(c_0) \) is compact, then there is a basis for \( \mathcal{N}[(vE)^T] \) for which all basis vectors have nonnegative integer entries.

If the reactions are partitioned into subsets of fast and slow steps we can write

\[
vE = v \left[ E^f \ | \ E^s \right]
\]

and it follows that any invariant of the full system is simultaneously an invariant of the fast and slow systems. We assume throughout that the slow and fast reactions are independent, and therefore \( \mathcal{R}(vE) = \mathcal{R}(vE^f) \cup \mathcal{R}(vE^s) \). Furthermore, one has that

\[
\mathcal{R}[vE^{(f,s)}] \subseteq \mathcal{R}[vE]
\]

and

\[
\mathcal{N}[(vE)^T] \subseteq \mathcal{N}[(vE^{(f,s)})^T].
\]

that the ranges of the slow and fast subsets are no larger than that of the full system, and the corresponding null spaces are no smaller, but the properties **P1, P2** of the full system do not necessarily carry over to the subsystems. However, it follows from (2.18) that one can define a map \( P^f : \mathbb{R}^m \rightarrow \mathbb{R}^{m-r_f} \) (where \( r_f = \text{dim} \mathcal{R}[vE^f] \)) for the fast subsystem that represents a vector in \( \mathcal{N}[(vE^f)^T] \) in terms of intrinsic coordinates on \( \mathcal{N}[(vE^f)^T] \). The associated matrix \( P^f \) has rows given by basis vectors with integer components of \( \mathcal{N}[(vE^f)^T] \). It follows that the reaction simplex for the fast subsystem is given by

\[
\Omega_f(c_0) \equiv \{ c \ : \ c \in c_0 + \mathcal{R}[vE^f] \} \cap \hat{\mathbb{R}}^+_m = \{ c : P^f c = P^f c_0 \equiv \tilde{c} \in \mathbb{R}^{m-r_f} \} \cap \hat{\mathbb{R}}^+_m.
\]

Here \( \tilde{c} \) represents a conserved quantity for the fast subsystem, but it may vary as slow reactions occur.

If the system is open then one can regard the effect of inputs and outputs as moving the dynamics between simplexes of fixed mass, as dictated by the evolution equations given at (2.9). This point of view will be useful in the stochastic formulation.
3 The stochastic formulation

3.1 The master equation

In a stochastic description the number of molecules of a species is too small to be treated as a continuous variable—they are random variables. We define \( N(t) = (n_1(t), n_2(t), \ldots, n_m(t)) \), where \( n_i(t) \) is as before, but now \( N(t) \) is a random vector. Under the assumption that the process is Markovian, which is appropriate if all the relevant species are taken into account, the evolution of \( N(t) \) is governed by a continuous-time Markov process with discrete states, and we denote the probability that \( \{N(t) = n\} \) by \( P(n, t) \). The governing equation for the evolution of \( P(n, t) \) is called the chemical master equation, and is given as

\[
\frac{d}{dt} P(n, t) = \sum_\ell R_\ell(n - v_\ell(n)) \cdot P(n - v_\ell(n), t) - \sum_\ell R_\ell(n) \cdot P(n, t) \tag{3.1}
\]

where \( E_\ell \), the \( \ell \)th column of \( E \), denotes \( \ell \)th reaction and the stochastic rates have the form

\[
R_\ell = c_\ell h_{j(\ell)}(n). \tag{3.2}
\]

Here \( c_\ell \) is the probability per unit time that the molecular species in the \( j \)th complex reacts, \( j(\ell) \) denotes the reactant complex for the \( \ell \)th reaction, and \( h_{j(\ell)}(n) \) is the number of independent combinations of the molecular components in this complex (Gadgil et al. 2005). One sees in (3.1) how the reaction graph \( G \) determines the state transition graph \( G_s \) that defines the steps in the master equation.

In the stochastic analysis all reactions are assumed to follow mass-action kinetics, and thus \( c_\ell = k_\ell \), and the combinatorial coefficient is given by\(^3\)

\[
h_{j(\ell)} = \prod_i \binom{n_i}{\nu_{ij(\ell)}}. \tag{3.3}
\]

If \( \nu_{ij(\ell)} = 1 \) the stochastic rate reduces to (2.13) but for a bimolecular reaction the rate of a step in the stochastic framework is always smaller than in the deterministic framework.

The master equation (3.1) can be expressed more explicitly for unimolecular and bimolecular mass action kinetic mechanisms. The general form for the uni- and bimolecular molecular reactions given in Table 1 is

\(^3\) This formulation applies only to ideal solutions—in nonideal solutions the number of molecules must be replaced by an appropriate measure of its activity in the solution. In particular, this involves a suitable description of diffusion when the solution is not ideal (Othmer 1976; Schnell and Turner 2004).
Table 1 The four classes of first-order reactions considered in (Gadgil et al. 2005) and two types of bimolecular reactions.

| Label | Type of reaction | Reaction | Rate |
|-------|-----------------|----------|------|
| I     | Production from a source | $\phi \rightarrow M_i$ | $k_s^i$ |
| II    | Degradation | $M_i \rightarrow \phi$ | $k_{d1}^i n_i$ |
| III   | Conversion | $M_j \rightarrow v_i M_i$ | $k_{con1}^i n_j$ |
| IV    | Catalytic production from source | $\phi M_j \rightarrow M_i$ | $k_{cat}^i j n_j$ |
| V     | Bimolecular degradation | $M_j + M_k \rightarrow \phi$ | $k_{d2}^i j n_j n_k$ |
| VI    | Bimolecular conversion | $M_j + M_k \rightarrow v_i M_i$ | $k_{con2}^i j n_k$ |

The relationship between the deterministic and stochastic rates of these reactions are discussed later. Rates are given in terms of number of molecules ($n$ will be introduced later)

$$\frac{dP(n, t)}{dt} = \sum_{i=1}^{s} \left[ K_{s}^i (S_i^{-1} - 1) P(n, t) + \sum_{j=1}^{s} \left( K_{con1}^i (S_i^{1} S_j^{-1} - 1) + \sum_{k=1}^{s} \left( K_{con2}^{ijk} (S_i^{1} S_j^{1} S_k^{-1} - 1) + K_{d2}^{ijk} (S_j^{1} S_k^{1} - 1) \right) n_k \right) (n_j P(n, t)) \right].$$

where $S_i^{k}$ is the shift operator that increases the $i$th component of $n$ by an integer amount $k$, and $S_i^{k} (n_i P(n, t)) = S_i^{k} n_i \cdot P(S_i^{k} n, t)$. In this expression the matrices $K_s^i$, $K_{d1}^i$, $K_{cat}^i$, $K_{d2}^i$, $K_{con1}^i$, and $K_{con2}^i$ correspond to the six types of reactions given in Table 1.

A compact representation of (3.1) is obtained by defining an ordering of the accessible states, of which we first assume there finitely many ($n_s$) in number and which lie in the non-negative cone $C_m^+ \subset R_n$. The evolution of the vector $p(t)$ of the probabilities of these states is governed by the matrix Kolmogorov equation

$$\frac{dp(t)}{dt} = K p(t), \tag{3.4}$$

where $K$ is the matrix of transition rates between the states, the entries of which are defined as follows. Let $n^i = (n_1^i, n_2^i, \ldots, n_m^i)$ and $n^j = (n_1^j, n_2^j, \ldots, n_m^j)$ denote the $i$th and $j$th states of the system and denote the $i$th and $j$th entries of the vector $p$ as $p_i = P(n^i, t)$ and $p_j = P(n^j, t)$, resp. Then the $(i, j)$th entry of $K$ is given by

$$K_{ij} = \begin{cases} \mathcal{R}_\ell(n^j) & \text{if } n^i = n^j + v \mathcal{E}(\ell) \text{ for some } \ell = 1, \ldots, r \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

The case in which there are infinitely many states will be discussed in Sect. 3.2.
and the following pseudo-code shows how to generate $K$ and the corresponding state transition graph $G_s$ for any reacting system. By starting with a given initial state vector, the algorithm checks all the reactions and adds resulting new state vectors into state space while updating the transition matrix, and then repeats the same procedure on the newly added state vector until no new state vector can be added.

---

**Data:** initial state vector $V_i$ and reactions $\{R_i\}_{i=1}^s$

**Result:** transition matrix $K_{s \times ns}$

Initialization: set current state index $C_s = 1$, set accessible state vector space $\{V\}$ to be $\{V_i\}$, set accessible state vector space size $ns = 1$, set transition rate matrix $K$ to be $\{0\}$;

while current state index $\leq ns$ do

set current reaction index to one: $C_r = 1$;

while current reaction $\leq r$ do

check if current reaction $R_{C_r}$ reacts from current state $V_C$;

if True then

get current state index: source $\leftarrow C_s$;

get target state $V_T = V_C + \alpha \epsilon_{(C_s,C_r)}$;

check if target state $V_T$ is already in $\{V\}$;

if True then

get the index $i$ of the state in $\{V\}$ that is equal to $V_T$;

check if $K_{i,\text{source}} = 0$;

if True then

update $K_{i,\text{source}}$;

end

end

else

add $V_T$ to $\{V\}$: $V_{ns} \leftarrow V_T$;

increase accessible state space size by one: $ns \leftarrow ns + 1$;

update $K_{ns,\text{source}}$;

end

end

increase current reaction index by one: $C_r \leftarrow C_r + 1$;

end

increase current state index by one: $C_s \leftarrow C_s + 1$;

end

update diagonal entries $K_{ij} = -\sum_{i=1,i\neq j}^{ns} K_{ij}$.

**Algorithm 1:** An algorithm for generating the transition matrix $K$ for a chemical reaction network

---

Off-diagonal elements in the $i$th row of $K$ are transition rates at which source states reach state $i$ in one step, whereas off-diagonal elements in the $i$th column represent the rates at which the $i$th state reaches its target states in one reaction. Since only uni- and bimolecular reactions are realistic, the states that reach the current state in one step (the sources) differ from the current state by at most two molecules, but those that can be reached in one step (the targets) may involve more than two. Under certain orderings $K$ has a narrow bandwidth when the number of states $ns$ is small, but expands with increasing $ns$. The number $ns$ grows combinatorially with the number of molecules—when there are $m$ species and a total of $N_0$ molecules

$$ns = \binom{N_0 + m - 1}{m - 1}.$$ 

If, for example, there are 4 species and 50 molecules, the number of states is 23,426.

In any case, if all reactive states are accounted for the matrix $K$ is the generator of a Markov chain (Norris 1998), i.e. $\sum_j K_{ij} = 0$ for each $j$, $K_{ij} \geq 0$ for each $j \neq i$, and the vector $1^T \equiv (1, 1, \ldots, 1)^T$ of length equal to $ns$ is in $N(K^T)$.

The formal solution of (3.4) is

$$p(t) = e^{Kt} p(0),$$
but there are well-known difficulties in computing the exponential if the state space
is large, and various approximate methods have been used (Kazeev et al. 2014; Menz
et al. 2012; Deuflhard et al. 2008). An alternate approach is to do direct stochastic
simulations, using the Gillespie algorithm or one of its many variants. However the
computational time can be extremely long if there are many species or the system is
spatially distributed (Hu et al. 2013), and thus it is advantageous to reduce the system
if possible.

Reduction of deterministic systems is frequently done by taking advantage of the
presence of multiple time scales in the evolution, and reducing the number of variables
by invoking the QSSH, as was done in (Lee and Othmer 2009). In the stochastic system
one can determine whether a reaction is fast or not according to the magnitude of the
transition rate of a reaction, the so-called propensity function in the chemical literature.
A larger transition rate implies that the corresponding step occurs more frequently, but
since the transition rate of a reaction depends on the numbers of reactant molecules as
well as the rate constant of the reaction, a reaction that is fast in the interior of $C^+_m$
may be slow near the boundary of the cone. The following example shows that a reaction
with a large rate constant may have to be considered as a slow reaction.

Example 1  Consider the following reaction network

$$A \xrightarrow{k_1} B \xleftarrow{k_{-1}}$$

If $k_1 = 0.01$, $k_{-1} = 0.1$, and initially $n_A(0) = 100$, $n_B(0) = 1$, then the rate of
$A \rightarrow B$ is $k_1n_A(0) = 1$, and from $B \rightarrow A$ is $k_{-1}n_B(0) = 0.1$. Thus even though
$k_1 < k_{-1}$, initially $A \rightarrow B$ is fast compared with $B \rightarrow A$. Of course when $A$ is small
$k_1n_A(t) < k_{-1}n_B(t)$, and which is fast and which is slow is interchanged.

This issue must be addressed on a case-by-case basis, but here we simply assume that
all steps can be identified as occurring on either an $O(1/\epsilon)$ scale or on an $O(1)$ scale
\textit{a priori}. Since we consider a small system with small numbers of molecular species
involved in at most second-order reactions, the separation assumption can be justified
by stipulating that

$$k_f \approx O\left(\frac{1}{\epsilon}\right) \quad \text{and} \quad N_0^2k_s \approx O(1),$$

where $N_0$ is the maximum number of molecules of reactants involved in slow reactions
and $k_f$ and $k_s$ are characteristic rate constants of fast and slow reactions, respectively.
Thus we can rewrite the matrix $K$ as

$$K = \frac{1}{\epsilon}K^f + K^s,$$

where $K^f$ and $K^s$ are the fast and slow transition matrices whose entries are the
transition rates of fast and slow reactions, respectively. Then on the $O(1)$ timescale
(3.4) can be written

\textcopyright Springer
\[
\frac{dp(t)}{dt} = \left(\frac{1}{\epsilon} K^f + K^s\right) p(t).
\] (3.6)

As before, conservation of probability must be satisfied, but now it must hold separately on the slow and fast scales, i.e., \(1^T \in \mathcal{N}((K^f)^T)\) and \(\mathcal{N}((K^s)^T)\). Thus both the transition rate matrices \(K^s\) and \(K^f\) for slow and fast reactions are generators of Markov chains. As we see later, the sub-graphs \(G^f_s\) and \(G^s_s\) of \(G_s\) associated with the fast and slow reactions may have several disjoint connected components when considered as undirected graphs. These need not be strong components of \(G_s\).

Equation (3.6) is written on a slow time, but on the fast timescale \(\tau = t/\epsilon\) it reads

\[
\frac{dp(t)}{d\tau} = (K^f + \epsilon K^s) p(t)
\] (3.7)

and from this one sees that the slow reactions act as a perturbation of the fast reactions on this scale. It follows that the perturbation can only have an effect on the dynamics represented by the eigenvector(s) corresponding to the zero eigenvalue of \(K^f\), and this will be exploited later.

### 3.2 Conditions under which the state space is bounded

A number of technical issues have to be discussed before proceeding further. Firstly, it is easy to see that the positive cone of \(\mathbb{Z}^n\) is invariant, which therefore preserves non-negativity of the probabilities. Furthermore, if the reaction simplex is compact there are only finitely many states in the system, and the generator of the Markov chain is a finite matrix. In that case the invariant distribution is well-defined and, given certain additional properties discussed later, it is unique. When the state space has infinitely many points the generator is an infinite matrix, and in fact, is unbounded as an operator on \(l_\infty\) because the entries of \(K\) depend on \(n\). In this case little is known in general about its spectral properties and invariant measure(s). This is in contrast with random walks on homogeneous lattices, where the generator is bounded even if the state space is infinite. Moreover, the fact that the corresponding deterministic description has a compact invariant set is of no import in the stochastic description, since large deviations from the mean dynamics are possible. However, as the following example shows, the probabilities of very large numbers may in general be very small under suitable conditions.

**Example 2** Consider the simple process \(\phi \xrightarrow{k_1} A \xrightarrow{k_2} \phi\), where as usual, \(\phi\) represents a source and sink, and let \(p_n(t)\) be the probability of having \(n\) molecules of \(A\) in the system at time \(t\). Then one can show that the generating function

\[
G(s, t) = \sum_{n=0}^{\infty} s^n p_n(t)
\]

is

\[\triangleleft\text{ Springer}\]
The solution of this for the initial condition $p_1(0) = 1, \ p_n(0) = 0$ otherwise, is

$$G(s, t) = \left( 1 + (s - 1)e^{-k_2 t} \right) \cdot \exp\left( \frac{k_1}{k_2} (s - 1)(1 - e^{-k_2 t}) \right)$$

and from this one finds, by expanding this as a series in $s$, that

$$p_n(t) = \frac{1}{k_2 n!} \left( \frac{k_1}{k_2} \right)^{n-1} \left( 1 - e^{-k_2 t} \right)^{n-1} \times \left( k_1(1 - e^{-k_2 t})^2 + k_2 n e^{-k_2 t} \right) \cdot \exp\left( -\frac{k_1}{k_2} (1 - e^{-k_2 t}) \right) \quad (3.8)$$

Therefore the stationary distribution is

$$\lim_{t \to \infty} p_n(t) = \frac{1}{n!} \left( \frac{k_1}{k_2} \right)^n \exp\left( -\frac{k_1}{k_2} \right) ,$$

which is a Poisson distribution with parameter $k_1/k_2$. Thus $p_n(t)$ is non-zero for arbitrarily large $n$ in both the transient and stationary distributions, but it decays rapidly with $n$. For example, if $k_1/k_2 \sim O(1)$ and $n = 25$, $p_n \sim O(10^{-25})$ in the stationary distribution. Even if the stationary mean $k_1/k_2 \sim O(10)$, $p_n \leq 10^{-20}$ for $n \geq 50$ (one must always choose $n$ greater than the mean in order that $p_k < p_n$ for $k > n$).

While this is only a heuristic justification for truncating the state space to a box of the form $\Pi_{i=1}^m [0, M_i]$, we shall do this in the remainder of the paper. One could formalize the truncation by modifying the dynamics so as to make all states that lie in the positive cone and outside the box transient, and then restricting the initial data to lie in the box. Further work on the validity of this truncation is needed, but in any case the following results apply rigorously to closed systems and those systems whose inputs are terminated when the total number of the molecules being input reaches a pre-determined level, as in the following example.

**Example 3** Consider a triangular system in which there is a input of one species, as shown below.

$$\phi \xrightarrow{k_0} A \xrightarrow{k_1}{/} B \xrightarrow{k_3} C \xrightarrow{k_5}{/} A.$$

If the initial state is $(n_A(0), n_B(0), n_C(0)) = (0, 0, 0)$ this defines the initial simplex by $n_A(0) + n_B(0) + n_C(0)$. Whenever a molecule of $A$ is added the simplex shifts upward by one unit in the first octant. To insure that the simplex remains bounded we simply turn off the input $\phi$ when the number of species $A$ achieves a certain threshold,
say, \( N_A \). At this moment \( t = t^* \), we also observe the molecular number of species B and C, which are \( n_B(t^*) \), \( n_C(t^*) \), and the system evolves on this reaction simplex thereafter, since the system is closed thereafter.

In general we shall assume that the inputs and outputs occur in the slow dynamics, and thus the fast reactions occur on a fixed finite simplex. The inputs and outputs can move the system ‘up’ and ‘down’ in the positive cone, but we assume that this occurs on the slow time scale and if the total inputs are controlled the state space will remain bounded.

### 3.3 The role of invariants

Any conservation condition that applies to a system defined in terms of concentrations applies when the system is defined in terms of numbers, and therefore the invariants that characterize (2.6) apply in the stochastic framework as well. In particular, the definition of a kinetic manifold \( \Omega_f \) for the fast subsystem, in which the slow variable \( \tilde{c} \) is invariant, carries over with only minor modification. Since the state space is discrete in a stochastic description, only the set of integer points in a reaction simplex \( \Omega \) are relevant, and we call the subset of integer points in \( \Omega \) the discrete reaction simplex and denote it by \( \mathcal{L} \).

For any \( c \in \Omega(c_0) \), we can write \( c = c_0 + \nu \mathcal{E} \eta \geq 0 \), where \( \eta \) is the extent, and therefore

\[
n = N_A V c = N_A V c_0 + N_A V \nu \mathcal{E} \eta \in \Omega(N_A V c_0) = \Omega(n_0).
\]

Since \( n \) lies in \( \mathbb{Z}_m^+ \), which is the closure of the set \( \mathbb{Z}_m^+ \) of \( m \)-dimensional vectors with positive integer entries, it follows that the set of all integer points in the reaction simplex is given

\[
\mathcal{L}(n_0) \equiv \Omega(n_0) \cap \mathbb{Z}_m^+,
\]

and we call this the (full) discrete reaction simplex through \( n_0 \). Analytically, a discrete simplex \( \mathcal{L}(n_0) \) is defined as a coset of \( \mathcal{R}(\nu \mathcal{E}) \), and it can be generated numerically using Algorithm 1. To illustrate this, consider the following reaction network for an open system.

**Example 4** Consider the network

\[
\phi \rightarrow A \rightleftharpoons B \rightarrow \phi.
\]

where the input and output reactions are slow steps. One finds that

\[
\mathcal{R}(\nu \mathcal{E}) \cap \mathbb{Z}_2^+ = \text{span} \left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} \cap \mathbb{Z}_2^+.
\]
which is $\mathbf{Z}_2^T$. If we start with the initial state $(1, 0)$, and stop the input reaction once one molecule of $A$ is added from the source, as described in Sect. 3.2, we can generate the full discrete reaction simplex as the set of all points denoted ‘•’ Fig. 1a.

Since the inputs and outputs occur on the slow time scale, these reactions are denoted by red arrows in Fig. 1b, and the remaining reactions are in fast components (green arrows). There are two distinct fast (absorbing) strong components, as shown in Fig. 1b, defined as $L_{f1} = \{(1, 0), (0, 1)\}$ and $L_{f2} = \{(2, 0), (1, 1), (0, 2)\}$. Since $\mathcal{R}(\nu E_f) = \text{span}\{[-1, 1]^T\}$, $\mathcal{N}[(\nu E_f)^T] = \text{span}\{[1, 1]\}$, we can identify each fast simplex uniquely by

$$\tilde{n} = A^f n = [1, 1] \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = [n_1 + n_2],$$

where $n_1(n_2)$ denotes the number of molecules of species $A$ ($B$). The rows of the matrix $A^f$ comprise a basis for $\mathcal{N}[(\nu E_f)^T]$, and thus $L_{f1}$ is defined by $\tilde{n} = 1$, and $L_{f2}$ by $\tilde{n} = 2$. In this simple example, we can interpret $\tilde{n}$ as the total mass, which is conserved by the fast reactions.

To define fast simplexes in general we consider only fast reactions, and first identify all distinct absorbing strong components in $\mathcal{G}_f$. Then, working backwards, we identify the sources and internal strong components that feed into an absorbing component. This identifies a sub-graph of $\mathcal{G}_s$ comprised of an absorbing strong component and its ‘feeders’, and leads to the following definition of a fast simplex.

**Definition 5** A fast discrete reaction simplex $L_{f}$ is the set of vertices in an absorbing strong component and its predecessor sources and internal strong components.

**Remark 6** 1. A fast simplex is uniquely determined by its absorbing strong component. Each fast simplex is characterized by the invariants of the fast dynamics. These are

$$\tilde{n} = A^f n$$

(3.9)
where the rows of the matrix $A^f$ comprise a basis for $\mathcal{N}[(\nu E^f)^T]$. 

2. The invariant distribution of the fast dynamics on a given fast simplex is non-zero only for the states in the corresponding absorbing component.

3. Both source components and internal strong components may belong to more than one fast simplex, as is exemplified by a reaction network with a binary tree structure. In such a network the root belongs to all fast simplexes and internal nodes at the $k$th level belong to $2^{n-k}$ simplexes, where $n$ is the depth of the tree.

4. If there are no slow reactions in the network there is only one fast simplex, and this case is not of interest here.

5. In general the disjoint union of all fast simplexes is the full fast simplex, i.e.

$$\cup \mathcal{L}^f \equiv \Omega^f (n_0) \cap \mathbb{Z}^+.$$ 

Although the following reaction network due to Wilhelm (2009) involves unrealistic trimolecular steps, we use it to illustrate how to define the fast simplexes when the fast subgraph $G^f_s$ has multiple absorbing components whose predecessors are not disjoint.

**Example 7**

$$S + 2X \xrightarrow{k_1} 3X$$
$$3X \xrightarrow{k_2} 2X + P$$
$$X \xrightarrow{k_3} P$$

The species are $S$, $X$, $P$, and the complexes: $S+2X$, $3X$, $2X+P$, $X$, $P$ are labeled 1-5 in this order. Then the matrices $\nu$, $E$ and $\nu E$ are as follows.

$$\nu = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 3 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \nu E = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & -1 \\ 0 & 1 & 1 \end{bmatrix}.$$ 

It follows that $\rho(E) = 3$, $\rho(\nu E) = 2$, and the deficiency, which is defined as $\delta = \rho(E) - \rho(\nu E) = 1$. Clearly $S+X+P$ is a constant and equivalently, $\mathcal{N}[(\nu E)^T] = \text{span} \left((1, 1, 1)^T\right)$. If we set the initial state as $(n_S(0), n_X(0), n_P(0)) = (1, 2, 0)$, then the possible transitions in the system are as shown in Fig. 2a.

Assume that only the second reaction is slow and the others are fast. Considering only the fast transitions indicated by green directed edges, it is clear that the resulting $G^f_s$ has the two absorbing strong components $(1, 0, 2)$ and $(0, 0, 3)$ because there are two possible exit steps from the node $(1, 2, 0)$, one with rate $k_1$, the other with rate $k_3$. The two fast simplexes are
Fig. 2 Example 7: a the full discrete reaction simplex (black circles); b, c the fast simplexes

\[ \mathcal{L}_1^f = \{ (1, 2, 0), (1, 1, 1), (1, 0, 2) \} \]

\[ \mathcal{L}_2^f = \{ (1, 2, 0), (0, 3, 0), (0, 2, 1), (0, 1, 2), (0, 0, 3) \} \]

as shown in Fig. 2b, c. Figuratively, and also in the language of graph theory, the simplexes are branches of the tree that defines the full simplex in Fig. 2a. \( \mathcal{L}_1^f \) lies on the coset of \( \mathcal{R}(v \mathcal{E}^f_1) = span\{[0 \ -1 \ 1]^T\} \), thus \( \mathcal{N}[(v \mathcal{E}^f_1)^T] = span\{[1 \ 0 \ 0], [0 \ 1 \ 1]\} \), and \( \mathcal{L}_1^f \) is identified by \( \bar{n}_1 = (1, 2) \). \( \mathcal{L}_2^f \) lies on the coset of \( \mathcal{R}(v \mathcal{E}^f_2) = span\{[-1 \ 1 \ 0]^T, [0 \ -1 \ 1]^T\} \), thus \( \mathcal{N}[(v \mathcal{E}^f_2)^T] = span\{[1 \ 1 \ 1]\} \), and \( \mathcal{L}_2^f \) is identified by \( \bar{n}_2 = 3 \).

In general a component of the reaction graph may have more than one absorbing component and more than one fast simplex, as in the foregoing example. To determine them in general one must first identify the number of strong components and their type in every component graph \( \mathcal{G}_\alpha \subset \mathcal{G} \). Since no vertex in a source is reachable from any vertex outside its component and no vertex in a sink is reachable from a vertex in any other sink, the relationship of reachability defines a partial order on the strong components of \( \mathcal{G}_\alpha \). This in turn leads to the acyclic skeleton \( \mathcal{G}_\alpha \), which is defined as follows. Associate a vertex \( \bar{V}_j \) with each strong component of \( \mathcal{G}_\alpha \), and introduce a directed edge from \( \bar{V}_i \) to \( \bar{V}_j \) if and only if one (and hence every) vertex in \( V_{\alpha i} \) is reachable from \( V_{\alpha j} \). \( \mathcal{G}_\alpha \) is connected since \( \mathcal{G}_\alpha \) is connected, but it is acyclic; in fact, it
is a directed tree. One of two cases obtains: either $\hat{G}_\alpha$ consists of a single vertex and no edges, which occurs when $\hat{G}_\alpha$ consists of one strong component, or it has at least one vertex of in-degree zero and at least one vertex of out-degree zero. One can relabel the strong components if necessary, and write the adjacency matrix of $\hat{G}_\alpha$ in the form

$$A = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
x & x & 0 & 0 & 0 \\
x & x & x & 0 & 0 \\
x & x & x & x & 0 \\
x & x & x & x & 0 \\
\end{bmatrix}$$

where the $x$’s represent blocks that may be non-zero. The diagonal blocks are square matrices of dimensions equal to the number of sources, the number of internal strong components, and the number of sinks, respectively. The vertices corresponding to internal strong components can always be ordered so that the central block is lower triangular, since the strong components are maximal with respect to inclusion of edges. The number of sinks or absorbing strong components is the number of zero columns of $A$.

4 The reduction of a master equation with two time scales

4.1 The splitting of the evolution equations

In this section we show how to obtain the lowest-order approximation to the slow dynamics for a system that is described by

$$\frac{dp}{dt} = \left( \frac{1}{\epsilon} K^f + K^s \right) p. \quad (4.1)$$

on the $O(1)$ timescale and by

$$\frac{dp}{d\tau} = (K^f + \epsilon K^s) p \quad (4.2)$$

on the fast time scale $\tau = t/\epsilon$.

By a suitable ordering of the states we can write the fast transition matrix $K^f$ as a block diagonal matrix.
wherein the number of blocks is equal to the number of fast components in the state transition graph $G_s^f$ of the fast reactions. We show later in Theorem 11 that each block has zero as a semisimple eigenvalue and we analyze the structure of the blocks in detail, but here we first define the index of a matrix (Campbell and Meyer 1991).

**Definition 8** Let $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear transformation. The index of $A$ is the smallest non-negative integer $k$ such that $\mathcal{R}(A^{k+1}) = \mathcal{R}(A^k)$.

An equivalent definition is that the index of $A$ is the largest Jordan block corresponding to the zero eigenvalue of $A$. It follows from the general theory in (Campbell and Meyer 1991) that a matrix with a semisimple zero eigenvalue has index 1, and as a result, the following properties hold.

(i) $\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A)$

(ii) $\mathcal{R}(A^2) = \mathcal{R}(A)$

Since we assume that the inputs and outputs only occur on the slow time scale, the fast dynamics occur on a fixed compact simplex in $\mathbb{C}^m$ and $K^f$ is a bounded operator that generates a Markov chain, since $1 \in \mathcal{N}(K^f)^T$. We define $\dim(\mathcal{N}(K^f)^T) = n_f$, and of course it follows that both $\mathcal{N}(K^f)$ and $\mathcal{N}((K^f)^T)$ have a basis of $n_f$ linearly-independent eigenvectors. Let $\Pi$ be the $n_s \times n_f$ matrix whose columns are eigenvectors corresponding to zero eigenvalues of $K^f$, and let $L$ be the $n_f \times n_s$ matrix whose rows are eigenvectors corresponding to zero eigenvalues of $(K^f)^T$. One can choose these to form a biorthogonal set and therefore $L \Pi = I_{n_f}$. Furthermore, $\Pi L = P_0$, where $P_0$ is the eigenprojection corresponding to the zero eigenvalue of $K^f$, and thus $\mathcal{R}(P_0) = \mathcal{N}(K^f)$ and $\mathcal{R}(P_0^T) = \mathcal{R}(I - P_0)$.

By property (i) above $\mathcal{R}(P_0) = \mathcal{N}(K^f)$ and to show that $\mathcal{R}(I - P_0) = \mathcal{R}(K^f)$ we observe that the adjoint $P_0^T$ leads to the decomposition $\mathcal{R}(P_0) = \mathcal{R}(P_0^T) \oplus \mathcal{R}(I - P_0^T)$. We have that $\mathcal{R}(P_0^T) = \mathcal{N}((K^f)^T)$ is orthogonal to $\mathcal{R}(I - P_0)$, and from the basic properties of projections and their adjoints (Kato 1966) it follows that $\mathcal{R}(I - P_0) = \mathcal{R}(K^f)$.

Since $\mathcal{R}(n_s) = \mathcal{N}(K^f) \oplus \mathcal{R}(K^f)$, we can write

$$p = P_0 p + (I - P_0) p = \Pi \tilde{p} + \Gamma \hat{p},$$

(4.4)

where $\tilde{p} = L p$ and $\Gamma$ is an $n_s \times (n_s - n_f)$ matrix whose columns are basis vectors of $\mathcal{R}(K^f)$. Then the matrix $[\Pi \mid \Gamma]$ is an $n_s \times n_s$ matrix whose columns are basis vectors of $\mathcal{R}(n_s)$. Using (4.4) in (4.1) leads to

---

6 This is the direct sum, but is generally not the orthogonal direct sum.
A multi-time-scale analysis of chemical reaction networks…

\[ \Pi \frac{d\tilde{p}}{dt} + \Gamma \frac{d\hat{p}}{dt} = \left( \frac{1}{\epsilon} K^f + K^s \right) (\Pi \tilde{p} + \Gamma \hat{p}). \]  

(4.5)

To obtain the evolution equations for \( \tilde{p} \) and \( \hat{p} \) separately, we define the \((n_s - n_f) \times n_s\) matrix \( V = [0 | I_{n_s-n_f}] [\Pi | \Gamma]^{-1} \), which has the property that \( V[\Pi | \Gamma] = [0 | I_{m-n_f}] \). In other words, \( V\Pi = 0, \ V\Gamma = I_{n_s-n_f} \), and by multiplying (4.5) by \( L \) and \( V \), respectively, we obtain

\[ \frac{d\tilde{p}}{dt} = L K^s (\Pi \tilde{p} + \Gamma \hat{p}), \]  

(4.6)

\[ \frac{d\hat{p}}{dt} = V \left( \frac{1}{\epsilon} K^f + K^s \right) (\Pi \tilde{p} + \Gamma \hat{p}) = \frac{1}{\epsilon} V K^f \Gamma \hat{p} + V K^s (\Pi \tilde{p} + \Gamma \hat{p}). \]  

(4.7)

where we used the fact that \( L\Gamma = 0 \) because \( \mathcal{N}((K^f)^T) \perp \mathcal{R}(K^f) \).

On the fast time scale \( \tau = t/\epsilon \) we obtain in the limit \( \epsilon \to 0 \) that

\[ \frac{d\tilde{p}}{d\tau} = 0, \]  

(4.8)

\[ \frac{d\hat{p}}{d\tau} = V K^f \Gamma \hat{p}. \]  

(4.9)

On this time scale the slow variable \( \tilde{p} \) remains constant while the fast variable \( \hat{p} \) evolves according to (4.9).

On the slow time scale

\[ \frac{d\tilde{p}}{dt} = L K^s (\Pi \tilde{p} + \Gamma \hat{p}), \]  

(4.10)

\[ \epsilon \frac{d\hat{p}}{dt} = (V K^f + \epsilon V K^s)(\Pi \tilde{p} + \Gamma \hat{p}). \]  

(4.11)

and the limit \( \epsilon \to 0 \) leads to the condition that

\[ V K^f \Gamma \hat{p} = 0. \]

**Lemma 9**

\[ V K^f \Gamma \hat{p} = 0 \]

*only if* \( \Gamma \hat{p} = 0 \).

**Proof** By definition \( \Gamma \hat{p} \in \mathcal{R}(K^f) \), and since \( \mathcal{R}((K^f)^2) = \mathcal{R}(K^f) \) by property (ii) above, it follows that \( K^f \Gamma \hat{p} = \Gamma q \) for some \( q \). Since \( V\Gamma = I_{n_s-n_f} \), \( V K^f \Gamma \hat{p} = 0 \) only if \( q = 0 \), which implies that \( \Gamma \hat{p} = 0 \).

Since \( \Gamma \hat{p} = 0 \), it follows from the Eq. (4.10) that

\[ \frac{d\tilde{p}}{dt} = L K^s \Pi \tilde{p} \equiv \tilde{K} \tilde{p}, \]  

(4.12)
Fig. 3  Triangular reaction: red and green arrows denote slow and fast reactions, respectively. 

a The reaction network; b the state transition diagram for a total of two molecules 

which shows how the slow evolution between fast simplexes depends on components of the fast evolution through $L$ and $\Pi$. The components of $\hat{p}$ represent the probabilities of aggregated states, and since $\Gamma \hat{p} = 0$ it follows from (2.14) that $p = \hat{\Pi} \hat{p}$, from which one obtains the probabilities of the individual states.

The following example illustrates the structure underlying the general reduction.

**Example 10** Consider a closed triangular reaction system with all reversible reactions, in which reactions 1 and 2 are fast, as shown in Fig. 3. The total number of molecules is conserved in a closed system of first order reactions of this type, and the transitions in state space are as illustrated in Fig. 3 when the total number of molecules is two.

The matrix $K_f$ is obtained by setting the rates of all slow steps to zero. The resulting graph has three connected components, $(2,0,0) \leftrightarrow (1,1,0) \leftrightarrow (0,2,0), \ (1,0,1) \leftrightarrow (0,1,1)$ and $(0,0,2)$, and this leads to the representation

$$K = \frac{1}{\epsilon} \left[ \begin{array}{ccc} K^f_1 & 0 & 0 \\ 0 & K^f_2 & 0 \\ 0 & 0 & K^f_3 \end{array} \right] + \left[ \begin{array}{ccc} K^s_{1,1} & K^s_{1,2} & K^s_{1,3} \\ K^s_{2,1} & K^s_{2,2} & K^s_{2,3} \\ K^s_{3,1} & K^s_{3,2} & K^s_{3,3} \end{array} \right], \quad (4.13)$$

where $K^f_i, i = 1, 2, 3$ is the $m_i \times m_i$ matrix of transition rates within each fast component, and $K^s_{i,j}$ is an $m_i \times m_j$ matrix of slow transition rates between fast components. In this example, the $K^f_i$ are

$$K^f_1 = \begin{bmatrix} -2k_1 & k_2 & 0 \\ 2k_1 & -(k_1 + k_2) & 2k_2 \\ 0 & k_1 & -2k_2 \end{bmatrix}, \quad K^f_2 = \begin{bmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{bmatrix}, \quad K^f_3 = 0,$$

Note that the vector $\mathbf{1}$ of the appropriate dimension is a left eigenvector of each $K^f_i$, and therefore probability is conserved on each fast component. In view of the block structure of $K^f$ in (4.13) zero is a semisimple eigenvalue of $K^f$.  

Springer
The sub-matrices of $K^s$ are

$$K^s_1 = \begin{bmatrix}
-2k_6 & 0 & 0 \\
0 & -(k_3 + k_6) & 0 \\
0 & 0 & -2k_3
\end{bmatrix},
K^s_2 = \begin{bmatrix}
-(k_4 + k_5 + k_6) & 0 \\
0 & -(k_3 + k_4 + k_5)
\end{bmatrix},
K^s_3 = -(2k_4 + 2k_5)$$

and

$$K^s_{2,1} = \begin{bmatrix}
2k_6 & k_3 & 0 \\
0 & k_6 & 2k_3
\end{bmatrix},
K^s_{1,2} = \begin{bmatrix}
k_5 & 0 \\
k_4 & k_5
\end{bmatrix},
K^s_{3,1} = \begin{bmatrix}
0 & 0 & 0
\end{bmatrix}^T,
K^s_{3,2} = \begin{bmatrix}
k_6 & k_3
\end{bmatrix},
K^s_{2,3} = \begin{bmatrix}
2k_5 & 2k_4
\end{bmatrix}^T.$$

The matrices $L$ and $\Pi$ are given by the following.

$$L = \begin{bmatrix}
L_1 & 0 & 0 \\
0 & L_2 & 0 \\
0 & 0 & L_3
\end{bmatrix},
\Pi = \begin{bmatrix}
\Pi_1 & 0 & 0 \\
0 & \Pi_2 & 0 \\
0 & 0 & \Pi_3
\end{bmatrix},$$

where $L_1 = [1 \ 1 \ 1]$, $L_2 = [1 \ 1]$, $L_3 = 1$ are basis vectors of $N(K^f)$’s. The $\Pi_i$’s are the multinomial invariant distributions on the fast components (Gadgil et al. 2005), and are found to be

$$\Pi_1 = \begin{bmatrix}
k_2^2 \\
2k_1k_2 \\
k_1^2
\end{bmatrix} \frac{1}{(k_1 + k_2)^2},
\Pi_2 = \begin{bmatrix}
k_2 \\
k_1 + k_2 \\
k_1 + k_2
\end{bmatrix},
\Pi_3 = 1.$$

0’s are zero row or column vectors with the appropriate dimensions.

As shown above, the dynamics on the slow time scale are governed by (4.12), and in this example the transition rate matrix for the slow system is given by

$$\tilde{K} = L K^s \Pi = \begin{bmatrix}
L_1 K^s_1 \Pi_1 & L_1 K^s_{1,2} \Pi_2 & L_1 K^s_{1,3} \Pi_3 \\
L_2 K^s_{2,1} \Pi_1 & L_2 K^s_{2,2} \Pi_2 & L_2 K^s_{2,3} \Pi_3 \\
L_3 K^s_{3,1} \Pi_1 & L_3 K^s_{3,2} \Pi_2 & L_3 K^s_{3,3} \Pi_3
\end{bmatrix} = \begin{bmatrix}
\tilde{k}^s_{1,1} & \tilde{k}^s_{1,2} & \tilde{k}^s_{1,3} \\
\tilde{k}^s_{2,1} & \tilde{k}^s_{2,2} & \tilde{k}^s_{2,3} \\
\tilde{k}^s_{3,1} & \tilde{k}^s_{3,2} & \tilde{k}^s_{3,3}
\end{bmatrix}$$

where the off-diagonal transition rates $\tilde{k}^s_{i,j}$ are given by

$$\tilde{k}^s_{1,2} = L_1 K^s_{1,2} \Pi_2, \quad \tilde{k}^s_{1,3} = \tilde{k}^s_{3,1} = 0, \quad \tilde{k}^s_{2,1} = L_2 K^s_{2,1} \Pi_1, \quad \tilde{k}^s_{2,3} = L_2 K^s_{2,3} \Pi_3, \quad \tilde{k}^s_{3,2} = L_3 K^s_{3,2} \Pi_2.$$

The diagonal elements are the negatives of the corresponding column sums and therefore $\tilde{K}$ is the generator of a Markov chain. We will show later—in Theorem 13—that...
the transition rate matrix for the reduced system is always a Markov chain generator. Recall as discussed in Sect. 4.3, we have

\[ E^f = \begin{bmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 0 \end{bmatrix} \]

\[ A^f = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

\[ \tilde{n} = A^f n = \begin{bmatrix} n_1 + n_2 \\ n_3 \end{bmatrix}, \]

Thus \( \tilde{n}_1 = (2, 0) \), \( \tilde{n}_2 = (1, 1) \), and \( \tilde{n}_3 = (0, 2) \). The evolution on the slow time scale is as shown schematically below.

\[ \begin{align*}
\kappa_{2,1}^t & \quad \kappa_{3,2}^t \\
(2, 0) \Leftrightarrow (1, 1) & \Leftrightarrow (0, 2) \\
\kappa_{1,2}^t & \quad \kappa_{2,3}^t
\end{align*} \]

The matrices for the general case in which the total number of molecules is \( N_0 \) are given in the Appendix.

### 4.2 The structure of the generator of the slow dynamics

We recall the standing assumption that there are no inputs or outputs that occur on the fast time scale, and therefore on the fast time scale the reactions are confined to fast simplexes as defined earlier. The slow dynamics moves the state between fast simplexes \( \Omega_f \) on a given simplex \( \Omega \), and/or between simplexes on the slow time scale. Because the transition matrix for the fast dynamics has the block diagonal form given in (4.3), both \( L \) and \( \Pi \) have the same block structure and it suffices to determine the null spaces for a fixed block.

The graph \( G_f^s \) can be decomposed into sources, internal strong components and absorbing strong components. If all fast reactions on fast simplexes are reversible, then all fast components are strongly connected. If a state or vertex is not connected to any other states, we call it an isolated state, and it is then a (absorbing) strong component itself. Note that in the triangular reaction network as in Example 10, each fast component is an absorbing strong component.

We let \( S_i \) be the set of all states (nodes) in the graph of the \( i \)th simplex and denote the set of the states in the sources, internal strong components and absorbing strong components by \( S_{so}^i \), \( S_{in}^i \) and \( S_{ab}^i \), respectively. The set of all the states, \( S \) can be represented by disjoint unions

\[ S = \dot{\bigcup}_i S_i = \dot{\bigcup}_i (S_{so}^i \cup S_{in}^i \cup S_{ab}^i). \]

We assume that \( |S_i| = m_i, |S_{so}^i| = m_{ab}^i, |S_{in}^i| = m_{in}^i \) and \( |S_{so}^i| = m_{so}^i \). Thus \( m_i = m_{ab}^i + m_{in}^i + m_{so}^i \).

By following the general analysis of reaction networks developed in (Othmer 1979) we can write each block \( K_i^f \) in (4.3) as an upper triangular block matrix
\[ K^f_i = \begin{bmatrix} K^f_{i,ab} & K^f_{i,ab,in} & K^f_{i,ab,so} \\ K^f_{i,ab,in} & K^f_{i,in} & K^f_{i,so,in} \\ K^f_{i,ab,so} & K^f_{i,so,in} & K^f_{i,so} \end{bmatrix}, \]

where each block \( K^f_{i,\alpha} \) is the transition matrix between states \( S_i^\alpha \), where \( \alpha = ab, in, so \) and \( K^f_{i,\alpha,\beta} \) is the transition matrix from the states in \( S_i^\beta \) into states in \( S_i^\alpha \), where \( \alpha, \beta = in, so \) or \( ab \).

If the \( i \)th simplex includes \( p_i \) isolated absorbing states and \( q_i \) absorbing strong components with at least two states, then the matrix \( K^f_{i,ab} \) can be written as

\[ K^f_{i,ab} = \begin{bmatrix} 0_{p_i} & K^f_{i,1,ab} & \cdots & K^f_{i,q_i,ab} \\ K^f_{i,1,1} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ K^f_{i,q_i,1} & \cdots & K^f_{i,q_i,q_i} \end{bmatrix}, \]

where \( 0_{p_i} \) is a \( p_i \times p_i \) zero matrix that reflects the transition rates between the \( p_i \) isolated absorbing states and \( K^f_{i,j} \) is an \( m_{i,j}^{ab} \times m_{i,j}^{ab} \) transition matrix describing the transitions between all states in the \( j \)th absorbing strong components with at least two states.

The following result defines an important spectral property of each \( K^f_i \), which determines the structure of its left and right eigenvectors \( L_i \) and \( \Pi_i \) as shown in Subsection 4.3.

**Theorem 11** Each matrix \( K^f_{i,ab} \) has a semisimple zero eigenvalue and each \( K^f_{i,in} \) and \( K^f_{i,so} \), \( i = 1, \ldots, l \), is nonsingular.

**Proof** The proof of this follows from the general result given in (Othmer 1979). This theorem implies that the zero eigenvalue of \( K^f_i \) is semisimple.

### 4.3 The invariant distributions of the fast dynamics

Next we consider the invariant distributions of the fast dynamics, which give the basis vectors in \( \Pi \). These are vectors \( \pi \geq 0 \) such that \( K^f \pi = 0 \) and \( \sum_j \pi_j = 1 \). Since every absorbing strong component with at least two states has a unique steady-state probability \( \pi_{i,j}^{ab}, j = 1, \ldots, q_i, \pi_{i,j}^{ab} \) is a basis for \( N(K^f_{i,j}) \) since \( K^f_{i,j} \pi_{i,j}^{ab} = 0 \), and \( \dim N(K^f_{i,j}) = 1 \). Notice that each absorbing strong component with only one state, which corresponds to each diagonal entry of the block \( 0_{p_i} \) in \( K^f_{i,ab} \), has steady-state probability 1. Using these facts we define
\[
\tilde{\Pi}_i = \begin{bmatrix} I_{p_i} & \pi_{i,1}^{ab} \\ \vdots & \ddots \\ \pi_{i,1}^{ab} & \cdots & \pi_{i,q_i}^{ab} \end{bmatrix},
\]

where \( I_{p_i} \) is an \( p_i \times p_i \) unit matrix and \( \pi_{i,j}^{ab} \) is an \( m_{i,j}^{ab} \times 1 \) vector. Since any states in sources and internal strong components have zero probability at the steady-state, we define the matrices

\[
\Pi_i = \begin{bmatrix} \tilde{\Pi}_i \\ 0_{r_i} & 0_{r_i} & \cdots & 0_{r_i} \\ 0_{s_i} & 0_{s_i} & \cdots & 0_{s_i} \end{bmatrix}, \quad \Pi = \begin{bmatrix} \Pi_1 \\ \vdots \\ \Pi_l \end{bmatrix},
\]

(4.14)

where \( 0_{r_i} \) and \( 0_{s_i} \) are null matrices representing the steady-state probability of all states in internal strong components and sources, respectively. Since \( K_f^i \Pi_i = 0 \) it follows that \( K_f^i \Pi = 0 \), and therefore we have an orthogonal basis for \( \mathcal{N}(K_f^i) \).

To construct a basis of \( \mathcal{N}(K_f^i)^T \), recall that the multiplicity of an eigenvalue is the same for a matrix and its adjoint, and therefore \( K_{i,j}^{f,ab} \) has exactly one zero eigenvalue and so \( \dim(\mathcal{N}(K_{i,j}^{f,ab})^T) = 1 \) for each \( j \). The \( 1 \times m_{i,j}^{ab} \) vector \( \tilde{L}_{i,j}^{ab} \equiv [1, \ldots, 1] \) is an eigenvector corresponding to the zero eigenvalue of \( N(K_{i,j}^{f,ab})^T \), and therefore is a basis of \( \mathcal{N}[(K_{i,j}^{f,ab})^T] \) for each \( j \). Define

\[
\tilde{L}_i = \begin{bmatrix} I_{p_i} \\ \tilde{L}_{i,1}^{ab} \\ \vdots \\ \tilde{L}_{i,q_i}^{ab} \end{bmatrix}
\]

and note that \( \tilde{L}_i K_i^{f,ab} = 0 \). By defining \( L_i = \begin{bmatrix} \tilde{L}_i & A_i & B_i \end{bmatrix} \), where \( A_i = -\tilde{L}_i K_i^{f,ab,\text{in}} (K_i^{f,\text{in}})^{-1} \) and \( B_i = -(\tilde{L}_i K_i^{f,ab,\text{so}} + A_i K_i^{f,\text{in,so}}) (K_i^{f,\text{so}})^{-1} \), one can see that

\[
L_i K_i^f = \begin{bmatrix} K_i^{f,ab} & K_i^{f,ab,\text{in}} & K_i^{f,ab,\text{so}} \\ K_i^{f,\text{in}} & K_i^{f,\text{in,so}} & K_i^{f,\text{so}} \end{bmatrix} = 0.
\]

Thus,

\[
LK^f = 0,
\]
where
\[
L \equiv \begin{bmatrix} L_1 & \cdots & L_l \end{bmatrix}.
\] (4.15)

One can see that all row vectors of \( L_i \) consist of a basis of \( N((K_f^T)^T) \) and it follows from the block diagonal structure of \( L \) that all row vectors of \( L \) are basis vectors of \( N((K_f^T)^T) \). One can also show that the sets \( \{L_i\} \) and \( \{\Pi_j\} \) are biorthogonal.

**Remark 12** When all components are strongly connected, e.g., all fast reactions are reversible, the matrices \( L \) and \( \Pi \) reduce to
\[
\begin{align*}
\tilde{L} = & \begin{bmatrix} \tilde{L}_1 & \cdots & \tilde{L}_l \end{bmatrix}, \\
\Pi = & \begin{bmatrix} \Pi_1 & \cdots & \Pi_l \end{bmatrix},
\end{align*}
\]
where \( \tilde{L}_i \) is a \( 1 \times m_i \) vector \([1, 1 \cdots 1]\) and \( \Pi_i \) is an \( m_i \times 1 \) steady-state vector of a strong component \( L_i \). Here \( m_i \) is the number of states in \( L_i \).

The following theorem shows that the slow evolution is always Markovian, even when a fast simplex has multiple components, each of which may have sources and/or internal components.

**Theorem 13** \( \tilde{K} = L K^s \Pi \) is the generator of a Markov chain.

**Proof** In order to prove this, we only need to show that the sum of each column of \( \tilde{K} \) is zero and the off-diagonal entries are nonnegative. Without loss of generality and to avoid complicated subindices, we consider the first column of \( \tilde{K} \). The first column block of \( \tilde{K} \), denoted as \( (\tilde{K})_1 \), is given by
\[
(\tilde{K})_1 = [L_1 K^s_{i,1} \Pi_1 \ L_2 K^s_{2,1} \Pi_1 \cdots \ L_n K^s_{n,1} \Pi_1]^T,
\]
where \( n \) is the number of fast components, block \( L_i K^s_{i,1} \Pi_1 \in \mathbb{R}^{ab_i \times ab_1}, \ i = 1, \ldots, n, \ ab_i \) is the number of absorbing strong components of the \( i \)-th fast component. The sum of all the rows of \( \tilde{K}_1 \) is
\[
\sum_i \left( \sum_{k_i} (L_i K^s_{i,1} \Pi_1)^\text{row } k_i \right) = \sum_i \sum_{k_i} (L_i K^s_{i,1})^\text{row } k_i \Pi_1 = \sum_i \left( \sum_{k_i} L_i^\text{row } k_i \right) K^s_{i,1} \Pi_1
\]
\[
= \sum_i 1_{1 \times D_i} K^s_{i,1} \Pi_1 = \sum_i 1_{1 \times D_i} K^s_{i,1} \Pi_1 = 0_{1 \times D_i} \Pi_1 = 0_{1 \times ab_1},
\]
where \( A^\text{row } k_i \) denotes the \( k_i \)-th row of matrix \( A \) and \( D_i \) is the number of nodes in the \( i \)-th fast component. The third equality follows from the fact that the column sum of \( L_i \) is always one, because each \( L_i \) defines an invariant probability and the sum must be
one. The last equality follows from the fact that the column sum of $K^s_{i,1}$ is zero. Since non-positive entries appear only along the diagonal of $K^s_{i,1}$, the off-diagonal entries of $\tilde{K}$ are nonnegative.

To summarize, we now have the approximate equation for the probability distribution on the slow time scale

$$\frac{d\tilde{p}}{dt} = LK^s\Pi\tilde{p},$$  \hspace{1cm} (4.16)

where the matrices $L$ and $\Pi$ are given above. Clearly the utility of the reduction depends heavily on whether the invariant distributions of the fast dynamics are easily computed. This is possible for a network of first-order reactions, and for closed systems the distribution is multinomial if the initial distribution is multinomial, while for open first-order systems the distribution is a product Poisson (Gadgil et al. 2005). The general case of first-order reactions is completely solved in (Jahnke and Huisinga 2007), where the authors show that the solution can be represented as the convolution of multinomial and product Poisson distributions with time-dependent parameters that evolve according to the standard equations.

Analytical results for bimolecular reactions are more limited. However it is known that the invariant distributions of a class of Markov chains that arise in queuing theory have a product form (Boucherie and Dijk 1991), and it has been shown that this also holds for a zero-deficiency reaction network (which means that $\mathcal{R}(\mathcal{E}) \cap \mathcal{N}(\nu) = \phi$) under the assumption of ideal mass-action kinetics (Anderson and Kurtz 2011; Mélykútí et al. 2014). When this does not apply the stationary distributions frequently involve hypergeometric functions (McQuarrie 1967).

### 4.4 A detailed example of the reduction

To illustrate the reduction method on general reaction networks, we consider the three reaction networks shown in Fig. 4. The resulting states beginning from one molecule of A and two of F are shown in Table 2, where the states are in the order of molecular species $A, B, \ldots, G$. The state space of network 3 is a subset of that of networks 1 and 2 since reactions with rates $k_3$ and $k_5$ are missing, as seen in see Table 2 (right). The corresponding full state networks as shown in Fig. 5 are generated using MATLAB.

By switching off the slow reactions in the graphs of the full state transition diagram shown in Fig. 5 we obtain three fast components for each network as shown in Fig. 6 (top), (center) and (bottom) respectively. The fast components on the right for every network are strongly connected and are the same since they are all generated by the reversible reactions $2F \leftrightarrow G$. By highlighting the nodes appearing in the same absorbing strong component, we see that in the state diagram of network 1, each fast component has a unique absorbing strong component. Comparing the top and center graphs in Fig. 6, we see the only difference occurs in the middle component. Network 2 has one more absorbing strong component than network 3, and the remainder are the same.

We perform the reduction method on the three networks and obtain the reduced slow transition matrices for each network as follows. The detailed steps in the reduction are
shown in supplemental material available at http://math.umn.edu/~othmer/Reduction.pdf. One finds that the columns of the $3 \times 3$ matrix $\tilde{K}_1$, the generator of the slow dynamics for network 1, are given by

\[
\begin{pmatrix}
-\tilde{k}_0 k_{11} (k_1 + k_2) & -\tilde{k}_0 k_{11} (k_1 + k_2) & \tilde{k}_0 k_{11} (k_1 + k_2) \\
\tilde{k}_0 k_{11} (k_1 + k_2) & \tilde{k}_0 k_{11} (k_1 + k_2) & \tilde{k}_0 k_{11} (k_1 + k_2) \\
\tilde{k}_0 k_{11} (k_1 + k_2) & \tilde{k}_0 k_{11} (k_1 + k_2) & \tilde{k}_0 k_{11} (k_1 + k_2)
\end{pmatrix}
\]

\[
\begin{pmatrix}
\tilde{k}_1 & \tilde{k}_1 & \tilde{k}_1 \\
\tilde{k}_1 & \tilde{k}_1 & \tilde{k}_1 \\
\tilde{k}_1 & \tilde{k}_1 & \tilde{k}_1
\end{pmatrix}
\]

\[
\begin{pmatrix}
\tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) & \tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) & \tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) \\
\tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) & \tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) & \tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) \\
\tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) & \tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7) & \tilde{k}_2 k_2 (k_2 + k_3 + k_4 + k_5 + k_6 + k_7)
\end{pmatrix}
\]

\[
\begin{pmatrix}
\tilde{k}_3 k_3 & \tilde{k}_3 k_3 & \tilde{k}_3 k_3 \\
\tilde{k}_3 k_3 & \tilde{k}_3 k_3 & \tilde{k}_3 k_3 \\
\tilde{k}_3 k_3 & \tilde{k}_3 k_3 & \tilde{k}_3 k_3
\end{pmatrix}
\]

\[
\begin{pmatrix}
6\tilde{k}_0 k_{11} (k_1 + k_2) & 3k_0 k_{10} (k_1 + k_2) & 0 \\
3k_0 k_{10} (k_1 + k_2) & 3k_0 k_{10} (k_1 + k_2) & -6k_0 k_{11} (k_1 + k_2) \\
0 & -6k_0 k_{11} (k_1 + k_2) & 3k_0 k_{10} (k_1 + k_2)
\end{pmatrix}
\]
Fig. 5 The graphs of the full state transition diagram with initial state \((1, 0, 0, 0, 2, 0)\) for network 1–3, red arrows denote slow reactions. a Network 1. b Network 2. c Network 3. Red and green arrows denote slow and fast reactions, respectively.

The first three columns of \(\tilde{K}_2\) are given by

\[
\begin{pmatrix}
-(k_3 + k_8) - \frac{k_9 k_{11}}{k_10 + k_{11}} + \frac{k_9 k_2}{k_2 + k_4} & \frac{k_5 k_7}{(k_2 + k_4)(k_6 + k_7)} & \frac{k_9 k_{11}}{k_10 + k_{11}} \\
\frac{k_3 k_4}{k_2 + k_4} & \left(-\frac{k_3 k_7}{(k_2 + k_4)(k_6 + k_7)}\right) & \frac{k_9 k_{11}}{k_10 + k_{11}} \\
0 & 0 & -(k_3 + 2k_8) + \frac{k_5 k_2}{k_2 + k_4}
\end{pmatrix},
\]
Fig. 6 The three fast components for network 1 (top), network 2 (center) and network 3 (bottom). The first has three fast components and three absorbing strong components, the second has three fast components and six absorbing strong components, and the third has three fast components and five absorbing strong components.
and the last three columns are given by

\[
\begin{pmatrix}
(K_2)_4 \\
0 \\
k_3 \\
\frac{-k_3k_7}{k_6+k_7} \\
\frac{k_6}{k_2+k_4} \\
0
\end{pmatrix}
\begin{pmatrix}
(K_2)_5 \\
0 \\
k_3k_7 \\
\frac{-k_3k_2k_7}{(k_2+k_4)(k_6+k_7)} \\
\frac{k_6}{k_2+k_4} \\
0
\end{pmatrix}
\begin{pmatrix}
(K_2)_6 \\
0 \\
0 \\
0 \\
0 \\
\frac{-6k_9k_{11}(k_{10}+k_{11})}{3k_{10}^2+6k_{10}k_{11}+k_{11}^2}
\end{pmatrix}
\]

Finally,

\[
\tilde{K}_3 = \begin{pmatrix}
-k_8 - \frac{k_9k_{11}}{k_{10}+k_{11}} & 0 & 2k_8 & 0 & \frac{6k_9k_{11}(k_{10}+k_{11})}{3k_{10}^2+6k_{10}k_{11}+k_{11}} \\
0 & -\frac{k_9k_{11}}{k_{10}+k_{11}} & 0 & k_8 & 0 \\
\frac{k_9k_{11}}{k_{10}+k_{11}} & 0 & -2k_8 & 0 & 0 \\
0 & \frac{k_9k_{11}}{k_{10}+k_{11}} & 0 & -k_8 & 0 \\
k_8 & 0 & 0 & 0 & -\frac{6k_9k_{11}(k_{10}+k_{11})}{3k_{10}^2+6k_{10}k_{11}+k_{11}}
\end{pmatrix}
\]

A number of general facts emerge from this example.

1. If each fast component has a unique absorbing strong component, entries of the left eigenvector of the fast transition matrix \( K_f \) are all one’s, i.e. \( L = [1, 1, \ldots, 1] \). In this case, there is no need to calculate \( A' \)’s and \( B' \)’s.
2. The column sum of \( L_i \) is always one.
3. The dimension of the reduced slow transition matrix \( \tilde{K} \) is the total number of absorbing strong components. Our reduction method does not require the uniqueness of absorbing strong component in each fast component. Without loss of generality, we consider the middle component of network 3 for example, see Fig. 7. It is obvious that the probability of \( \odot \) choosing to transit to \( \odot \) is \( k_2/(k_2+k_4) \), and the probability of transiting to \( \odot \) is \( k_4/(k_2+k_4) \). Conceptually we can suppose that this 5-state component is comprised of \( \odot \rightarrow \odot 
\rightarrow \odot \) with probability \( k_2/(k_2+k_4) \), and occurs as \( \odot \rightarrow \odot \rightarrow \odot \rightarrow \odot \) with probability \( k_4/(k_2+k_4) \), and the reduction method reflects this correctly.

To see this, denote the probability vector corresponding to the 5-state component by

\[
P = [P_{11}, P_{14}, P_{19}, P_7, P_3]^T.
\]

Then the left eigenvector \( L \) is found to be

\[
L = \begin{bmatrix}
k_2 \\
k_2+k_4 \\
k_2+k_4 \\
k_2 \\
k_2+k_4
\end{bmatrix}
\]
and the probability vector of the reduced network is given by

$$\tilde{P} = LP = \begin{bmatrix} P_{11} + \frac{k_2}{k_2+k_4}(P_7 + P_3) \\ P_{14} + P_{19} + \frac{k_4}{k_2+k_4}(P_7 + P_3) \end{bmatrix},$$

the first row of $\tilde{P}$ corresponds to the probability of being trapped in the absorbing strong component $11$, whereas the second row corresponds to the probability of evolving to the other absorbing strong component $14 \xrightarrow{k_6} 19$. The accumulated state associated with $3 \xrightarrow{k_4} 7 \xrightarrow{k_4} 11$ is obtained by switching off the reaction with rate $k_4$, which is $(2, 0, 0)$, and similarly $3 \xrightarrow{k_4} 7 \xrightarrow{k_4} 14 \xrightarrow{k_6} 19$ gives rise to $(2, 1, 0)$.

4. Comparing the reaction network 2 and 3 as shown in Fig. 4, we can see network 2 has two more reverse slow reactions, i.e. reactions with rate $k_3$ and $k_5$. Consider the reduced slow transition matrices $\tilde{K}_2$ and $\tilde{K}_3$, if we let $k_3$ and $k_5$ be zeros in $\tilde{K}_2$, then $\tilde{K}_2$ and $\tilde{K}_3$ are the same except $\tilde{K}_2$ has one extra row and column of zeros, which correspond to the absorbing strong component $25 \xrightarrow{k_6} 26 \xrightarrow{k_7} 27$ that only exists in network 2.

4.5 The mean and variance

Often the first and second moments of the distribution of each species is desired, and from the previous analysis one can find an approximate expression for the mean and variance of each species as follows. Define a projection operator $T_k$ as

$$T_k(n) = n_k,$$

where $n = (n_1, \ldots, n_s)$. Observing from the previous analysis that each state $S_{ij}$, the $j$th state in $\mathcal{L}_i$, has a probability distribution $\pi_{ij} \tilde{p}_i(t)$ at time $t$, one can find for each $k = 1, \ldots, s$
\[ E[n_k, t] = \sum_{i,j} T_k(n^{ij}) \hat{p}_i(t) \pi_{ij}, \quad (4.17) \]

\[ \text{Var}[n_k, t] = \sum_{i,j} [T_k(n^{ij})]^2 \hat{p}_i(t) \pi_{ij} - E[n_k, t]^2, \quad (4.18) \]

where \( n^{ij} \) denotes the vector \( n \) of molecular numbers corresponding to the state \( S_{ij} \). Thus we can obtain formal solutions (4.17) and (4.18) for mean and variance of each species on the slow dynamics.

In linear models, it is possible to obtain explicit expressions for mean and variance (Gadgil et al. 2005). In that case one can consider transitions between species as random walks of a molecule. This fact leads to a simple stochastic algorithm for computation and especially, if each fast component \( C_i \) in the reaction network is strongly connected, the quasi-steady-state distribution of the fast component is multinomial (Gadgil et al. 2005). Thus one can compute the conditional expectation \( E[R_s(n) | \tilde{n}] \) as follows; the quasi-steady-state probability vector \( \pi_i \) of a component \( C_i \) can be computed by using

\[ K^i_f \pi_i = 0, \quad \sum_{j=1}^{m_i} \pi_{ij} = 1, \]

where \( \pi_{ij} \) is a steady-state probability of \( j \)th species \( M_{ij} \) in the \( i \)th component. If the reactant of a slow reaction is \( M_{ij} \), then one can obtain the transition rate for the reaction

\[ E[R_s(n) | \tilde{n}] = K^s \pi_{ij} \tilde{n}, \]

where \( k_e \) is the transition rate constant of the reaction \( R^s \).

Further insight into the structure of the slow dynamics is gotten from the fact that entry of \((LK^s \Pi)\) is a conditional expectation when all fast components are strongly connected, as proven next.

**Theorem 14** Suppose that fast components are strongly connected. Then the approximate transition rate \( \tilde{R}_{\ell} \) for the transition rate of all slow reactions from \( j \)th to \( i \)th fast component is

\[ \tilde{R}_{\ell} = \sum_{\ell} c_{\ell} E[h_{\ell}(n) | \tilde{n}] = \sum_{\ell} E[R^s_{\ell}(n) | \tilde{n}], \]

where \( c_{\ell} \) is the transition rate constant of the \( \ell \)th slow reaction which transits states from \( j \)th to \( i \)th fast component and \( h_{\ell} \) is the combinatorial number of reactants of the \( \ell \)th slow reaction.

**Proof** First note that by an occurrence of the \( \ell \)th slow reaction, a state in the fast component \( L^f_j \) is transformed into a state in the fast component \( L^f_i \). When a fast component is strong the variable \( \tilde{n} = A_f n \) is uniquely defined in each component, we
suppose that \( \tilde{n} = \tilde{n}^i \) for \( L^i_f \) and \( \tilde{n} = \tilde{n}^j \) for \( L^j_f \). Let \( n^{i,k} \) be the vector of molecular numbers that denote the state \( S_{ik} \), the \( k \)th state in the component \( L^i_f \).

Since \( L K^s \Pi \) is a Markov chain generator, it follows from the reduced equation (4.12) that the approximate transition rate for a transition from the \( j \)th to the \( i \)th fast component is given by the \((i, j)\)th entry of \( L K^s \Pi \), which is computed as follows.

\[
(L K^s \Pi)_{ij} = (\text{diag}[L_1, \ldots, L_{n_f}] K^s \text{diag}[\Pi_1, \Pi_2, \ldots, \Pi_{n_f}])_{ij}
\]

\[
= [1 \ 1 \ \ldots \ 1] \begin{bmatrix}
(K^s_{ij})_{11} & \cdots & (K^s_{ij})_{1m_j} \\
(K^s_{ij})_{21} & \cdots & (K^s_{ij})_{2m_j} \\
\vdots & \ddots & \vdots \\
(K^s_{ij})_{m_1} & \cdots & (K^s_{ij})_{m_m_j}
\end{bmatrix}
\begin{bmatrix}
\pi_{j1} \\
\pi_{j2} \\
\vdots \\
\pi_{jm_j}
\end{bmatrix}
\]

\[
= \sum_q \sum_p (K^s_{ij})_{pq} \pi_{jq} = \sum_q c_\ell h_\ell(n^{i,q}) P(n = n^{i,q} | \tilde{n} = \tilde{n}^j) = \sum_q c_\ell [R^s_\ell(n) | \tilde{n} = \tilde{n}^j] = E[R^s_\ell(n) | \tilde{n} = \tilde{n}^j]
\]

where the \((i, j)\)th block matrix of \( K^s \), \( K^s_{ij} \), is a matrix of transition rates from the states of \( L^j_f \) to those of \( L^i_f \) and so \((K^s_{ij})_{pq}\) is the rate of transitions from the \( q \)th state in \( L^j_f \) into the \( p \)th state in \( L^i_f \). In the fourth equality, we made use of the fact that \( \sum_p (K^s_{ij})_{pq} = \sum_\ell c_\ell h_\ell(n^{i,q}) \) and \( \pi_{jq} \) is the conditional probability \( P(n = n^{i,q} | \tilde{n} = \tilde{n}^j) \).

**Remark 15** The result is true in general but the proof is somewhat more involved since the structure of the \( L_i \) and the \( \Pi_j \) is more complicated.

When \( \tilde{n} \) defined earlier is uniquely defined in each discrete simplex \( L^j_f \), we have the relationship between the full and reduced systems shown in the following table.

| Variable          | Original system | Reduced system |
|-------------------|-----------------|----------------|
| Variable          | \( n \)         | \( \tilde{n} = A^f n \) |
| Stoichiometry     | \( v E^s \)     | \( v E^s = A^f v E^s \) |
| Transition rate   | \( R^s(n) \)    | \( \tilde{R}^s(n) = E[\sum R^s[\tilde{n}] \] |

Moreover, under the assumptions in Theorem 14 we can obtain an approximate chemical master equation on the slow time scale when the fast components are strongly connected. If the original master equation is given by

\[
\frac{d}{dt} P(n, t) = \sum_\ell \frac{1}{\epsilon} \left[ R^f_\ell(n - v E^f_{(\ell)}) \cdot P(n - v E^f_{(\ell)}, t) - R^f_\ell(n) \cdot P(n, t) \right] + \sum_k \left[ R^s_k(n - v E^s_{(k)}) \cdot P(n - v E^s_{(k)}, t) - R^s_k(n) \cdot P(n, t) \right],
\]
then the master equation in the reduced system is approximated (with error less than $O(\epsilon)$)

$$\frac{d \tilde{p}(\tilde{n}, t)}{dt} = \sum_{k} \tilde{R}^s_k(\tilde{n} - \nu \tilde{E}^s_k) \tilde{p}(\tilde{n} - \nu \tilde{E}^s_k, t) - \tilde{R}^s_k(\tilde{n}) \tilde{p}(\tilde{n}, t), \quad (4.19)$$

where $\tilde{n} = A f \tilde{n}$, $\nu \tilde{E}^s(k) = A f \nu \tilde{E}^s(k)$ and $\tilde{R}^s_k(\tilde{n}) = E[\sum R^s_k(n)|\tilde{n}]$. In a later section we obtain a modified stochastic simulation algorithm from (4.19). If there are sources or internal strong components the master equation for the slow system contains additional terms.

5 A stochastic simulation algorithm for the slow dynamics

5.1 A stochastic simulation algorithm based on the QSS approximation

Preparation:

- Identify fast and slow reactions.
- For a given initial state $n(0)$, switch off the slow reactions and generate the transition matrix $K^f$ of the fast component that $n(0)$ lies in by implementing Algorithm 1.
- Identify the fast simplex that $n(0)$ lies in to determine $A^{f,ab}$.
- Define $\tilde{n} = A^{f,ab} n$, $\nu \tilde{E}^s \equiv A^{f,ab} \nu \tilde{E}^s$.
- Denote the fast simplex that $n(0)$ lies in to be $\tilde{n}(0) = A^{f,ab} n(0) \equiv \tilde{n}_j$, i.e., $j$th fast simplex and its target fast simplexes through slow reactions, denoted as $\tilde{n}_i$(could be more than one).
- Identify the fast transition matrices within the $i$th and $j$th fast simplexes, respectively, denoted as $K^{f,ab}_i, K^{f,ab}_j$. Identify slow transition matrices from $j$th to $i$th fast simplexes, denoted as $K^s_{i,j}$.
- Compute $L_i$ and $\Pi_j$ such that $L_i K^{f,ab}_i = 0$ and $K^{f,ab}_j \Pi_j = 0$.

Step 1. Scheme for a fast simplex: Computing slow reaction rates from the QSS state of fast a simplex

1. Compute the slow transition rate $\tilde{R}^s(\tilde{n}_j) \equiv \tilde{K}^s_{i,j} = L_i K^s_{i,j} \Pi_j$ from the $j$th fast simplex to all its target simplexes.
2. Compute $\tilde{R}^s_{tot} = \sum \tilde{R}^s_{i}(\tilde{n}_j)$.

Step 2. Scheme for simulation of slow reactions: Simulating the slow reactions with the reaction rates obtained in Step 1

1. Generate two random numbers $r_1$ and $r_2$ from the uniform distribution on $(0, 1)$.
2. Set $\tau = -\frac{\log(r_1)}{\tilde{R}^s_{tot}}$ and choose $k$ such that $\sum_{i=1}^{k-1} \tilde{R}^s_{i}(\tilde{n}_j) < r_2 \tilde{R}^s_{tot} \leq \sum_{i=1}^{k} \tilde{R}^s_{i}(\tilde{n}_j)$.

Step 3. Update

Let $t \leftarrow t + \tau$ and $\tilde{n} \leftarrow \tilde{n} + \nu \tilde{E}^s_k$. Go to Step 1.
Remark 16 1. The main idea of the modified Gillespie algorithm is to perform the traditional algorithm on a reduced rather than the original reaction network. For instance, in Example 10,

\[ A^f = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \tilde{n}(0) = A^f n(0) = (2, 0) \quad \tilde{\nu}E^s = A^f \nu E^s = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \]

and \( \tilde{\nu}E^s \) provides the stoichiometric matrix for a reduced reaction network. The reduced reaction network can be written as

\[ \mathcal{M}_1 \rightleftharpoons \mathcal{M}_2, \]

where \( \mathcal{M}_1, \mathcal{M}_2 \) are what we call “pseudo species” in the reduced network, and the new initial state \( \tilde{n}(0) = (2, 0) \) corresponds to only two molecules of species \( \mathcal{M}_1 \). Each pseudo species represents one fast (absorbing) strong component, in this case

\[ \mathcal{M}_1 : A \rightleftharpoons B \quad \mathcal{M}_2 : C. \]

The slow reaction rates in the reduced network are computed using \( LK^s \Pi \) as before except \( K_f \) and \( K_s \) are for the original reaction diagram rather than state diagram, i.e.

\[ K^f_1 = A \begin{bmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{bmatrix} \quad K^f_2 = C \begin{bmatrix} 0 \end{bmatrix} \quad K^s_{2,1} = C \begin{bmatrix} k_6 & k_3 \end{bmatrix} \quad K^s_{1,2} = A \begin{bmatrix} k_5 \\ k_4 \end{bmatrix}. \]

After finding the forward reaction rate from \( \mathcal{M}_1 \) to \( \mathcal{M}_2 \) to be \( (k_1 k_3 + k_2 k_6)/(k_4 + k_5) \) and the backward rate \( k_4 + k_5 \), one can initialize the Gillespie algorithm as usual.

Example 10 is the special case in which each fast component is strongly connected, i.e. only fast absorbing strong components or sinks exist. In general, one should identify sink, internal, and source strong components in the reaction graph. Once they are identified, the graph becomes a tree. Then one can use for example depth-first search to identify the fast discrete simplexes as defined in 5 in order to find \( A^f \).

The algorithm entails computation of the matrix of the slow dynamics, and whether one does this once for all initially, which may be advantageous when doing multiple realizations, or on the fly when only a few realizations are desired, is a matter of choice. In single realizations the entire state space generated by Algorithm 1 may not be explored and ‘on-the-fly’ computation of the slow transition rates may be advantageous.

6 Applications

In this section we analyze three examples: a bacterial motor model, an enzyme-inhibitor model and a model from PFK system. These examples illustrate the reduction
when there are relatively few states in the system. In these cases one can find the approximate probability distribution of the slow variables by solving the reduced matrix equation $d\tilde{p}/dt = LK^s\Pi \tilde{p}$ directly. Of course the main step is to find the matrix in the reduced equation

$$
\begin{align*}
\frac{d\tilde{p}}{dt} &= LK^s\Pi \tilde{p}.
\end{align*}
$$

**Example 17** (A bacterial motor model)

This example arises as a model for control of the rotational bias in the flagellar motor of *E. coli* (Othmer 2005). At the base of the motor are sites at which the protein CheY$_p$ can bind, and the occupancy of the sites biases the probability of switching the direction of rotation of the motor. Here we consider the following scheme for binding CheY$_p$ (represented by $Y$) to these sites.

\[
\begin{array}{c}
\text{CW}_0 \leftrightarrow CW_1 \leftrightarrow CW_2 \cdots \leftrightarrow CW_{n-1} \leftrightarrow CW_n \\
\end{array}
\]

\[
\begin{array}{c}
\alpha_0 \downarrow \beta_0 \quad \alpha_1 \downarrow \beta_1 \quad \alpha_2 \downarrow \beta_2 \quad \cdots \quad \alpha_{n-1} \downarrow \beta_{n-1} \quad \alpha_n \downarrow \beta_n \\
\end{array}
\]

\[
\begin{array}{c}
\text{CCW}_0 \leftrightarrow CCW_1 \leftrightarrow CCW_2 \cdots \leftrightarrow CCW_{n-1} \leftrightarrow CCW_n \\
\end{array}
\]

\[
\begin{array}{c}
k_1 Y \\
(n-1)k_1 Y \\
2k_1 Y \\
k_1 Y \\
\end{array}
\]

\[
\begin{array}{c}
k_3 Y \\
(n-1)k_3 Y \\
2k_3 Y \\
k_3 Y \\
\end{array}
\]

Here $CW_k$ and $CCW_k$ represent clockwise and counterclockwise flagellar rotation, respectively. We assume that the horizontal transitions are fast, while the vertical transitions are slow, which leads to two strongly connected fast components comprising the horizontal steps.

Thus the two $(n + 1) \times (n + 1)$ fast reaction rate matrices are given by

\[
K_1^f = \begin{bmatrix}
-nk_1 Y & k_{-1} & 0 & \cdots \\
\frac{nk_1 Y}{k_{-1} + (n-1)k_1} & \cdot & \cdot & \cdot \\
(n-1)k_1 Y & \cdot & \cdot & \cdot \\
k_1 Y & nk_{-1} & \cdot & \cdot \\
\end{bmatrix}
\]

\[
K_2^f = \begin{bmatrix}
-nk_3 Y & k_{-3} & 0 & \cdots \\
\frac{nk_3 Y}{k_{-3} + (n-1)k_3} & \cdot & \cdot & \cdot \\
(n-1)k_3 Y & \cdot & \cdot & \cdot \\
k_3 Y & nk_{-3} & \cdot & \cdot \\
\end{bmatrix}
\]
Moreover, the slow reaction rate matrix is

\[ K^s = \begin{bmatrix} -A & B \\ A & -B \end{bmatrix}, \]

where

\[ A = \text{diag}(\alpha_0, \ldots, \alpha_n) \text{ and } B = \text{diag}(\beta_0, \ldots, \beta_n). \]

Let \( \Pi = [\Pi_1 \mid \Pi_2] \) and let \( \Pi_i, i = 1, 2 \) be stationary distributions of \( K^f_i, i = 1, 2 \) respectively, i.e., \( \Pi_i \) is the right eigenvector of \( K^f_i \) corresponding to the zero eigenvalue with \( \sum_i \Pi_{ij} = 1 \), where \( \Pi_{ij} \) denotes the \( j \)th entry of the vector \( \Pi_i \). Then the reduced equation can be written

\[ \frac{d\tilde{p}}{dt} = LK^s \Pi \tilde{p}, \]

where

\[ LK^s \Pi = \begin{bmatrix} -\sum_{i=1}^{n+1} \alpha_{i-1} \Pi_{1i} & \sum_{i=1}^{n+1} \beta_{i-1} \Pi_{2i} \\ \sum_{i=1}^{n+1} \alpha_{i-1} \Pi_{1i} & -\sum_{i=1}^{n+1} \beta_{i-1} \Pi_{2i} \end{bmatrix} \]

Thus we reduce the system into two-state model,

\[ CCW \xleftrightarrow{k^+} CW, \]

where \( k^+ = \sum_{i=1}^{n+1} \alpha_{i-1} \Pi_{1i} \) and \( k^- = \sum_{i=1}^{n+1} \beta_{i-1} \Pi_{2i} \). Here we note that \( k^\pm \) are functions of \( k_i, n, \alpha, \beta \) and \( Y \). These parameters are reported in the experimental literature.

Figure 8 illustrate the evolution of probabilities of \( CW \) and \( CCW \) obtained from the full and the reduced equations.

**Example 18** (An enzyme inhibition model) We consider an enzyme-substrate reaction system with a competitive inhibition

\[ E + S \xrightleftharpoons[k_1]{k_2} ES \rightarrow E + P, \quad E + I \xrightleftharpoons[k_4]{k_5} EI, \quad (6.1) \]

where \( E, S, I \) and \( P \) denote enzyme, substrate, inhibitor and product, respectively. Let \( n_i(t), i = 1, \ldots, 6 \) denote the number of molecules of \( E, S, ES, I, EI \) and \( P \), respectively. We assume that the two reversible reactions \( E+S \xrightleftharpoons{} ES, \quad E+I \xrightleftharpoons{} EI \) are much faster than the irreversible reaction \( ES \rightarrow E + P \). After finding
Fig. 8 Shown here are the time evolutions of the probability that the motor rotates clockwise \( Pr(CW) = 1 \) and counterclockwise \( Pr(CCW) = 1 \) with different initial conditions using \( k_1 = 2, k_{-1} = 1, k_3 = 2, k_{-3} = 1, \alpha_l = 0.1, \beta_l = 0.08, Y = 100 \). In the upper two figures, the initial condition is that the motor is rotating clockwise with no \( Y \) bound. In the lower two figures, the initial condition is that the motor is rotating counterclockwise with no \( Y \) bound. Blue lines are generated by computing the full system \( Pr(CW) = \sum Pr(CW_i), Pr(CCW) = \sum Pr(CCW_i) \), whereas red circles are for the reduced system

\[
A^f = \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix},
\]

one can identify the slow variable

\[
\tilde{n} = A^f n = (n_1 + n_3 + n_5, n_2 + n_3, n_4 + n_5, n_6)^T.
\]

Since the fast subsystem \( E + S \xleftrightarrow{\gamma} ES, \ E + I \xleftrightarrow{\gamma} EI \) has a deficiency of zero and is weakly reversible, one can use the result by (Anderson and Kurtz 2011) for finding the equilibrium probability of the fast subsystem; For the convenience of computation, we assume that \( k_1 = k_2 \) and \( k_3 = k_4 \). If the deterministic equilibrium values of \( E, S, ES, I \) and \( EI \) for the fast dynamics are denoted by \( c_1, c_2, c_3, c_4 \) and \( c_5 \), respectively, one can find the only one solution
\[ c_3 = \frac{\alpha_2(1 + \alpha_1 + \alpha_2 + \sqrt{4\alpha_1 + (1 - \alpha_1 + \alpha_2 + \alpha_3)^2})}{2(\alpha_2 + \alpha_3)}, \]
\[ c_4 = \frac{\alpha_3(1 + \alpha_1 + \alpha_2 + \sqrt{4\alpha_1 + (1 - \alpha_1 + \alpha_2 + \alpha_3)^2})}{2(\alpha_2 + \alpha_3)}, \]
and
\[ c_1 = \alpha_1 - c_3 - c_5, \quad c_2 = \alpha_2 - c_3, \quad c_4 = \alpha_3 - c_5, \]
where \( \alpha_1 = c_1 + c_3 + c_5, \alpha_2 = c_2 + c_3, \) and \( \alpha_3 = c_4 + c_5 \) are conserved quantities in the fast subsystem. Using the result of (Anderson et al. 2010), one finds the equilibrium probability of the fast subsystem as
\[ p(n_1, \ldots, n_5) = M \prod_{i=1}^{5} \frac{c_i^{n_i}}{n_i!}, \quad n_i = 0, 1, 2, \ldots \]
where \( M \) is the normalizing constant for \( \sum_n p(n) = 1 \) and \( n_1, \ldots, n_5 \) satisfy the conserved quantities \( n_1 + n_3 + n_5, n_2 + n_3 \) and \( n_4 + n_5 \). The following figure shows the simulation results for the slow variable \( \tilde{n}_4 \) (the number of product molecules) obtained from the approximate algorithm and the exact algorithm (Fig. 9).

Example 19 (A model for the PFK system) We consider a reaction network from a model of the PFK step in glycolysis (Othmer and Aldridge 1978)

\[ A_1 + E_1 \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} E_1A_1 \overset{k_2}{\rightarrow} E_1 + A_2 \]

Fig. 9 Enzyme-substrate model with an inhibitor. Comparison of approximate stochastic simulation algorithm (red dotted) to exact stochastic simulation algorithm (blue solid). Evolution of means and standard deviations of numbers of the product \( P \) when the initial condition \((E, S, ES, I, EI, P) = (5, 100, 0, 5, 0, 0)\) and \( k_1 = k_2 = k_4 = k_5 = 10, k_3 = 0.1 \). The results are based on 5000 realizations. Concerning the relative CPU time for one realization of the stochastic simulation, if the approximate algorithm takes 1 s, the exact algorithm takes about 3.6 s, when a quad-core machine with Windows 8.1 and MATLAB 2014 is used.
Here $A_1, A_2, E_1, E_1^*$ and $E_2$ denote $F6P$, $ADP$, the low activity and activated forms of free PFK and the enzyme for the ADP sink reaction, respectively, and $E_1A_1, E_1^*A_1$ and $E_2A_2$ represent enzyme-substrate complexes.

If we assume three binding/unbinding reactions are much faster than others, there are two different fast subsystems $C_1$ and $C_2$ in the reaction network,

$$C_1 : A_1 + E_1 \xrightarrow{k_1} E_1^*A_1, \quad A_1 + E_1^* \xleftarrow{k_3} E_1^*A_1$$

and

$$C_2 : A_2 + E_2 \xrightarrow{k_5} E_2A_2.$$  

First we can find stationary distribution of $C_2$ using the hypergeometric functions (McQuarrie 1967). Let initial numbers of $A_2$ and $E_2$ be $a_0$ and $b_0$, respectively and define $Q = \frac{k_{-1}}{k_1}$. If $b_0 \geq a_0$, then the stationary marginal distribution of $A_2$ is given by

$$P_{A_2}(k) = D \frac{Q^k (a_0 + c_0)(a_0 + c_0 - 1) \cdots (a_0 + c_0 - k + 1)}{k! (b_0 - a_0 + 1)(b_0 - a_0 + 2) \cdots (b_0 - a_0 + k)},$$

where $k = 1, \ldots, a_0 + c_0$ and $D$ is a normalization constant.

Similarly, if $a_0 > b_0$, then the stationary marginal distribution of $E_2$

$$P_{E_2}(k) = D \frac{Q^k (b_0 + c_0)(b_0 + c_0 - 1) \cdots (b_0 + c_0 - k + 1)}{k! (a_0 - b_0 + 1)(a_0 - b_0 + 2) \cdots (a_0 - b_0 + k)},$$

where $k = 1, \ldots, b_0 + c_0$.

Since the fast subsystem $C_1$ has a deficiency of zero and is weakly reversible, we can find the equilibrium probability similar to the enzyme inhibitor model; If we assume that $k_1 = k_2$ and $k_3 = k_4$ and the deterministic equilibrium values of $A_1, E_1, E_1A_1, E_1^*, E_1^*A_1$ are denoted by $c_i, i = 1, \ldots, 5$, respectively, then we can find the equilibrium probability of the fast subsystem $C_1$

$$p(n_1, \ldots, n_5) = M \prod_{i=1}^{5} c_i^{n_i} n_i!, \quad n_i = 0, 1, 2, \ldots$$

where $n_i, i = 1, \ldots, 5$ are the number of $A_1, E_1, E_1A_1, E_1^*, E_1^*A_1$, subject to the conserved quantities $n_1 + n_3 + n_5, n_2 + n_3$ and $n_4 + n_5$, $M$ is the normalizing constant.
A comparison of the approximate stochastic simulation algorithm (red dotted) to the exact stochastic simulation algorithm (blue solid) for the model of PFK reaction system. Evolution of means and standard deviations of numbers of the product when the initial condition \((A_1, E_1, A_1, E_1^*, A_1, A_2, E_2, E_2 A_2, P) = (100, 5, 0, 5, 0, 100, 5, 0, 0)\) and \(k_2 = k_4 = k_6 = 0.1, k_1 = k_{-1} = k_3 = k_{-3} = k_5 = k_{-5} = 10\). The results are based on 5000 realizations. Concerning the relative CPU time for one realization of the stochastic simulation, if the approximate algorithm takes 1 s, the exact algorithm takes about 2.9 s, when a quad-core machine with Windows 8.1 and MATLAB 2014 is used.

$$\sum_n p(n) = 1$$ and \(n_1 \ldots, n_5\) satisfy the conserved quantities \(n_1 + n_3 + n_5, n_2 + n_3\) and \(n_4 + n_5\).

\[
\begin{align*}
c_3 &= \frac{\alpha_2(1 + \alpha_1 + \alpha_2 + \alpha_3 + \sqrt{4\alpha_1 + (1 - \alpha_1 + \alpha_2 + \alpha_3)^2})}{2(\alpha_2 + \alpha_3)}, \\
c_4 &= \frac{\alpha_3(1 + \alpha_1 + \alpha_2 + \alpha_3 + \sqrt{4\alpha_1 + (1 - \alpha_1 + \alpha_2 + \alpha_3)^2})}{2(\alpha_2 + \alpha_3)}, \\
c_1 &= \alpha_1 - c_3 - c_5, \ c_2 = \alpha_2 - c_3, \ c_4 = \alpha_3 - c_5,
\end{align*}
\]

and \(\alpha_1 = c_1 + c_3 + c_5, \ \alpha_2 = c_2 + c_3\) and \(\alpha_3 = c_4 + c_5\) are conserved quantities in the fast subsystem.

The following figure shows the simulation results for the number of product, which is a slow variable (Fig. 10).

7 Conclusion

We developed a reduction method for stochastic biochemical reaction networks with coupled fast and slow reactions, and formulated an associated extension of the Gillespie method. The reduced equation for the time-dependent probability distribution on the slow time scale involves a projection in state space, and we show that the generator of the reduced system is also Markovian. We found that the transition rate for a reaction on the slow time scale is approximated by the expectation of the original transition rate conditional on the invariant distribution of the fast dynamics. Throughout the reduction the invariants of fast subsystems play a significant role, similar to what was found earlier in the analogous reduction of deterministic systems (Lee and Othmer, 2013).
When the stationary distribution of the fast dynamics can be computed, these can be used directly in a modified stochastic simulation algorithm. Several biological examples, including a model of motor behavior for a single flagellum, an enzyme-inhibitor model and a model for the PFK system, illustrate the numerical accuracy of the approximation for two lowest moments.

8 Appendix

8.1 The general case in Example 10

When there are $N_0$ molecules, there are $N_0 + 1$ fast components and the transition matrix $K$ is block tridiagonal, i.e.

$$K = \begin{bmatrix}
    -\frac{1}{\epsilon}K^f_1 + K^s_1 & K^s_1 & K^s_{1,2} \\
    K^s_{2,1} & -\frac{1}{\epsilon}K^f_2 + K^s_2 & K^s_{2,3} \\
    \vdots & \ddots & \ddots \\
    K^s_{N_0, N_0+1} & \frac{1}{\epsilon}K^f_{N_0+1, N_0} + K^s_{N_0+1} & \ldots
\end{bmatrix},$$

where the fast blocks are given by

$$K^f_1 = \begin{bmatrix}
    -N_0k_1 & k_2 \\
    N_0k_1 & -(N_0 - 1)k_1 + k_2 & 2k_2 \\
    (N_0 - 1)k_1 & -(N_0 - 2)k_1 + 2k_2 & 3k_2 \\
    \vdots & \ddots & \ddots \\
    k_1 & -(N_0 - 1)k_2 & \ldots \\
\end{bmatrix}_{(N_0+1) \times (N_0+1)}$$

$$K^f_2 = \begin{bmatrix}
    -\frac{(N_0 - 1)k_1}{k_2} & k_2 \\
    \frac{(N_0 - 1)k_1}{k_2} & -(N_0 - 2)k_1 + k_2 & 2k_2 \\
    \frac{(N_0 - 2)k_1}{k_2} & -(N_0 - 3)k_1 + 2k_2 & 3k_2 \\
    \vdots & \ddots & \ddots \\
    k_1 & -(N_0 - 1)k_2 & \ldots \\
\end{bmatrix}_{N_0 \times N_0}$$

$$K^f_{N_0} = \begin{bmatrix}
    -k_1 & k_2 \\
    k_1 & -k_2
\end{bmatrix}, \quad K^f_{N_0+1} = 0.$$
The slow off-diagonal blocks are given by

\[
K_{1,2}^s = \begin{bmatrix}
    k_5 & k_4 & k_5 \\
    & \ddots & \ddots \\
    \ddots & k_5 & k_4 \\
\end{bmatrix}_{(N_0+1) \times N_0}, \quad K_{2,3}^s = \begin{bmatrix}
    2k_5 & 2k_4 & 2k_5 \\
    & \ddots & \ddots \\
    \ddots & 2k_5 & 2k_4 \\
\end{bmatrix}_{N_0 \times (N_0-1)}, \\
\] 

\[
K_{N_0,N_0+1}^s = \begin{bmatrix}
    N_0k_5 \\
    N_0k_4 \\
\end{bmatrix}.
\]

For the lower diagonal blocks,

\[
K_{2,1}^s = \begin{bmatrix}
    N_0k_6 & k_3 \\
    (N_0-1)k_6 & 2k_3 \\
    & \ddots & \ddots \\
    \ddots & k_6 & N_0k_3 \\
\end{bmatrix}_{N_0 \times (N_0+1)}, \quad K_{3,2}^s = \begin{bmatrix}
    (N_0-1)k_6 & k_3 \\
    (N_0-2)k_6 & 2k_3 \\
    & \ddots & \ddots \\
    \ddots & k_6 & (N_0-1)k_3 \\
\end{bmatrix}_{(N_0-1) \times N_0}, \ldots
\]

\[K_{N_0+1,N_0}^s = \begin{bmatrix}
    k_6 & k_3 \\
\end{bmatrix}.\]

Finally, the \(K_i^s\) along the diagonal are diagonal matrices of the same dimension as \(K_i^f\), and the \((j,j)\)-th entry of \(K_i^s\) is the negative sum of rates leaving \(j\)-th node of the \(i\)-th fast component.

In this case the transition rate is given by

\[
\tilde{k}_{i,j}^s = L_i K_{i,j}^s \pi_j.
\]

### 8.2 Moment equations of the invariant distributions

The low-order moments of the distributions for the fast systems play a role in the QSS reduction in Sect. 4, and here we consider the low-order moment equations.

**Theorem 20** Let \(r\) be the total number of the reactions in the system. Then the invariant (steady-state) distribution of \(P(n, t)\), which we denote \(P(n)\), satisfies

\[
\sum_{\ell=1}^r R_{\ell}(n - vE(\ell)) P(n - vE(\ell)) = \sum_{\ell=1}^r R_{\ell}(n) P(n) \quad (8.1)
\]
and the first two moment equations lead to

\[ E[\nu_\mathcal{E}\mathcal{R}(n)] = 0, \]
\[ \sum_{i=1}^{r} [\nu_\mathcal{E}(i) \otimes E[n\mathcal{R}_i(n)] + E[n\mathcal{R}_i(n)] \otimes \nu_\mathcal{E}(i) + \nu_\mathcal{E}(i) \otimes \nu_\mathcal{E}(i) E[\mathcal{R}_i(n)]] = 0. \]  

(8.2) \hspace{1cm} (8.3)

Proof At the steady-state

\[ \sum_{\ell=1}^{r} \mathcal{R}_\ell(n - \nu_\mathcal{E}(\ell)) P(n - \nu_\mathcal{E}(\ell)) = \sum_{\ell=1}^{r} \mathcal{R}_\ell(n) P(n) \]  

(8.4)

By multiplying by \( n \) and summing over all the values of \( n \in \mathcal{L}(n_0) \), we obtain

\[ \sum_{n} \sum_{\ell=1}^{r} n\mathcal{R}_\ell(n - \nu_\mathcal{E}(\ell)) P(n - \nu_\mathcal{E}(\ell)) = \sum_{n} \sum_{\ell=1}^{r} n\mathcal{R}_\ell(n) P(n). \]

Using the transformation \( n - \nu_\mathcal{E}(\ell) \to n \) on the left side, we obtain

\[ \sum_{n} \sum_{\ell=1}^{r} (n + \nu_\mathcal{E}(\ell)) \mathcal{R}_\ell(n) P(n) = \sum_{n} \sum_{\ell=1}^{r} n\mathcal{R}_\ell(n) P(n). \]

By subtracting the right side from the left one,

\[ \sum_{\ell=1}^{r} \nu_\mathcal{E}(\ell) \sum_{n} \mathcal{R}_\ell(n) P(n) = \sum_{\ell=1}^{r} \nu_\mathcal{E}(\ell) E[\mathcal{R}_\ell(n)] = 0. \]

Thus we conclude that

\[ \nu_\mathcal{E} E[\mathcal{R}(n)] = 0. \]

If the deficiency \( \delta \equiv \rho(\mathcal{E}) - \rho(\nu_\mathcal{E}) \) is zero, then \( E[\mathcal{R}(n)] \) is a cycle in the graph (Othmer 1979). \(^7\)

\(^7\) The reader can show that the presence of inputs or outputs of the form given in Table 1 does not alter the deficiency.
At the next order we multiply Eq. (8.4) by a tensor product $n \otimes n$ and sum over $n$. Then by a similar argument, we obtain

$$
\sum_{i=1}^{r} [v \mathcal{E}_{(i)} \otimes \sum_{n} n \mathcal{R}_{i}(n) p(n) + \sum_{n} n \mathcal{R}_{i}(n) p(n) \otimes v \mathcal{E}_{(i)}] + v \mathcal{E}_{(i)} \otimes v \mathcal{E}_{(i)} \sum_{n} \mathcal{R}_{i}(n) p(n)
$$

$$
= \sum_{i=1}^{r} [v \mathcal{E}_{(i)} \otimes E[n \mathcal{R}_{i}(n)] + E[n \mathcal{R}_{i}(n)] \otimes v \mathcal{E}_{(i)} + v \mathcal{E}_{(i)} \otimes v \mathcal{E}_{(i)} E[\mathcal{R}_{i}(n)]]
$$

$$
= 0.
$$

If $\delta = 0$, the two lowest moment equations can be simplified to

$$
E[\mathcal{R}(n)] = 0 \quad (8.5)
$$

$$
\sum_{i=1}^{r} [v \mathcal{E}_{(i)} \otimes E[n \mathcal{R}_{i}(n)] + E[n \mathcal{R}_{i}(n)] \otimes v \mathcal{E}_{(i)}] = 0. \quad (8.6)
$$

When all reactions are linear the problem is much simpler, and the evolution equations for the first and second moments can be written explicitly in terms of those moments (Gadgil et al. 2005).

As a consequence of Theorem 20, similar equations can be obtained for the quasi-steady-state of the probability distribution for the fast subsystem in a two-time scale stochastic network. We first define the expectation of a function $f(n)$ over a discrete reaction simplex $\mathcal{L}_{f}$ for the fast subsystem as follows.

$$
E_{\mathcal{L}_{f}}[f(n)] \equiv \sum_{n \in \mathcal{L}_{f}} f(n) p(n).
$$

**Corollary 21** Let $r_{f}$ be the total number of fast reactions. Then at the steady-state of the fast subsystem, the governing equation is given by

$$
\sum_{i=1}^{r_{f}} \mathcal{R}_{i}^{f}(n - v \mathcal{E}_{(i)}^{f}) P(n - v \mathcal{E}_{(i)}^{f}) = \sum_{i=1}^{r_{f}} \mathcal{R}_{i}^{f}(n) P(n) \quad (8.7)
$$

and for each discrete reaction simplex $\mathcal{L}_{f}$,

$$
E_{\mathcal{L}_{f}}[\mathcal{R}^{f}(n)] = 0, \quad (8.8)
$$

$$
\sum_{i=1}^{r_{f}} [v \mathcal{E}_{(i)}^{f} \otimes E_{\mathcal{L}_{f}}[n \mathcal{R}_{i}^{f}(n)] + E_{\mathcal{L}_{f}}[n \mathcal{R}_{i}^{f}(n)] \otimes v \mathcal{E}_{(i)}^{f}] = 0. \quad (8.9)
$$
Proof The ‘state-wise’ form of the master Eq. (3.1) can be written
\[
\frac{d}{dt} P(n, t) = \sum_\ell \mathcal{R}_f^\ell (n - v \mathcal{E}_f(\ell)) \cdot P(n - v \mathcal{E}_f(\ell), t) - \mathcal{R}_f^\ell(n) \cdot P(n, t)
\]
+ \sum_k \left[ \mathcal{R}_s^k(n - v \mathcal{E}_s(k)) \cdot P(n - v \mathcal{E}_s(k), t) - \mathcal{R}_s^k(n) \cdot P(n, t) \right],
\]
where $\mathcal{R}_f$ and $\mathcal{R}_s$ are the transition rates of fast and slow reactions, respectively and $\mathcal{E}_f$ and $\mathcal{E}_s$ are incidence matrices for fast and slow reactions, respectively.

In the previous theorem, substitute $\mathcal{E}_f$, $\mathcal{R}_f$ and $\mathcal{E}_f$ into $\mathcal{E}$, $\mathcal{R}$ and $E[\cdot]$ and use the full rank assumption on $v \mathcal{E}_f$.

References

Anderson DF, Craciun G, Kurtz TG (2010) Product-form stationary distributions for deficiency zero chemical reaction networks. Bull Math Biol 72:1947–1970
Anderson DF, Kurtz TG (2011) Continuous time markov chain models for chemical reaction networks. In: Koeppl H, Densmore D, Setti G, di Bernardo M (eds) Design and analysis of biomolecular circuits: engineering approaches to systems and synthetic biology. Springer, New York, pp 3–42
Aris R (1965) Prolegomena to the rational analysis of chemical reactions. Arch Ration Mech Anal 19(2):81–99
Boucherie RJ, Dijk NMV (1991) Product forms for queueing networks with state-dependent multiple job transitions. Adv Appl Probab 23(1):152–187
Bundschuh R, Hayot F, Jayaprakash C (2003) Fluctuations and slow variables in genetic networks. Biophys J 84:1606
Campbell SL, Meyer CP (1991) Generalized inverses of linear transformations. Dover Publications, New York
Cao Y, Gillespie DT, Petzold LR (2005) The slow-scale stochastic simulation algorithm. J Chem Phys 122:014116
Chen WK (1971) Applied graph theory. North-Holland, Amsterdam
Chevalier MW, El-Samad H (2009) A rigorous framework for multiscale simulation of stochastic cellular networks. J Chem Phys 131(5):054102
Cotter S (2015) Constrained approximation of effective generators for multiscale stochastic reaction networks and application to conditioned path sampling. arXiv:1506.02446 (arXiv preprint)
Deuflhard P, Huisinga W, Jahnke T, Wulkow M (2008) Adaptive discrete Galerkin methods applied to the chemical master equation. SIAM J Sci Comput 30(6):2990–3011
E W, Liu D, Vanden-Eijnden E (2005) Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. J Chem Phys 123:194107
Garidel C, Lee CH, Othmer HG (2005) A stochastic analysis of first-order reaction networks. Bull Math Biol 67:901–946
Gillespie DT (2007) Stochastic simulation of chemical kinetics. Annu Rev Phys Chem 58:35–55
Goutsias J (2005) Quasi-equilibrium approximation of fast reaction kinetics in stochastic biochemical systems. J Phys Chem 112:184102
Goutsias J, Jenkinson G (2013) Markovian dynamics on complex reaction networks. Phys Rep 529:199–264
Haseltine EL, Rawlings JB (2002) Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J Chem Phys 117(15):6959–6969
Hellander A, Löststedt P (2007) Hybrid method for the chemical master equation. J Comput Phys 227:100–122
Horn F, Jackson R (1972) General mass action kinetics. Arch Ration Mech Anal 48:81
Hu J, Kang H-W, Othmer HG (2013) Stochastic analysis of reaction–diffusion processes. Bull Math Biol 76:854–894
Huang C, Liu D (2014) Strong convergence and speed up of nested stochastic simulation algorithm. Commun Comput Phys 15(4):1207–1236
Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. J Math Biol 54:1–26
Janssen JAM (1989a) The elimination of fast variables in complex chemical reactions II. J Stat Phys 57:171–186
Janssen JAM (1989b) The elimination of fast variables in complex chemical reactions. III. Mesoscopic level (irreducible case). J Stat Phys 57:187–198
Kato T (1966) Perturbation theory for linear operators. Springer-Verlag, Berlin
Kazeev V, Khammash M, Nip M, Schwab C (2014) Direct solution of the chemical master equation using quantized tensor trains. PLoS Comput Biol 10(3):e1003359
Kim JK, Josic K, Bennett MR (2014) The validity of quasi-steady-state approximations in discrete stochastic simulations. Biophys J 107:783–793
Lee CH, Lui R (2009) A reduction method for multiple time scale stochastic reaction networks. J Math Chem 46:1292–1321
Lee CH, Othmer HG (2009) A multi-time-scale analysis of chemical reaction networks: I. Deterministic systems. J Math Biol 60(3):387–450
Mastny EA, Haseltine EL, Rawlings JB (2007) Two classes of quasi-steady-state model reductions for stochastic kinetics. J Chem Phys 127(9):094106
McQuarrie DA (1967) Stochastic approach to chemical kinetics. J Appl Probab 4(3):413–478
Mélykúti B, Hespanha JP, Khammash M (2014) Equilibrium distributions of simple biochemical reaction systems for time-scale separation in stochastic reaction networks. J R Soc Interface 11(97):20140054
Menz S, Latorre JC, Schutte C, Huisinga W (2012) Hybrid stochastic-deterministic solution of the chemical master equation. Multiscale Model Simul 10(4):1232–1262
Norris JR (1998) Markov chains. Cambridge University Press, Cambridge
Othmer HG (1976) Nonuniqueness of equilibria in closed reacting systems. Chemical Eng Sci 31:993–1003
Othmer HG (1979) A graph-theoretic analysis of chemical reaction networks Lecture Notes, Rutgers University. http://math.umn.edu/~othmer/graphrt.pdf
Othmer HG (2005) Analysis of complex reaction networks Lecture Notes, University of Minnesota
Othmer HG, Aldridge JA (1978) The effects of cell density and metabolite flux on cellular dynamics. J Math Biol 5:169–200
Peles S, Minsky B, Khammash M (2006) Reduction and solution of the chemical master equation using time scale separation and finite state projection. J Chem Phys 125:204104
Rao CV, Arkin AP (2003) Stochastic chemical kinetics and the quasi-steady state assumption: application to the Gillespie algorithm. J Chem Phys 118(11):4999–5010
Salis H, Kaznessis Y (2005) Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. J Chem Phys 122:054103
Schnell S, Turner T (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. Prog Biophys Mol Biol 85(2):235–260
Smith S, Cianci C, Grima R (2015) Model reduction for stochastic chemical systems with abundant species. arXiv:1510.03172 (arXiv preprint)
Srivastava R, Haseltine EL, Mastny E, Rawlings JB (2011) The stochastic quasi-steady-state assumption: reducing the model but not the noise. J Chem Phys 134(15):154109
Thomas P, Straube AV, Grima R (2011) Communication: limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks. J Chem Phys 135(18):181103
Wilhelm T (2009) The smallest chemical reaction system with bistability. BMC Syst Biol 3(1):90