# Error bound for hybrid models of two-scaled stochastic reaction systems

Tobias Jahnke and Vikram Sunkara

## 1 Introduction

Biological systems such as gene-regulatory networks and cell metabolic processes consist of multiple species which are undergoing transformations via a set of reaction channels. If all populations are sufficiently large, then the evolution of the concentrations over time can be modeled by the classical reaction-rate equation, i.e. a system of ordinary differential equations; cf. [23]. In many applications, however, some of the species occur in low amounts, and it was observed that small stochastic fluctuations in their populations can cascade large effects to the other species. Important examples are gene-regulatory networks where the evolution of the entire system depends crucially on the stochastic behavior of a rather small number of transcription factors. In order to capture these effects, such systems must be described by a Markov jump processes, which respects the inherent discrete nature of the system and its stochastic interactions.

The associated time-dependent probability distribution is the solution of the *Chemical Master Equation (CME)*, but solving the CME is a considerable challenge, as the size of the state space scales exponentially in the number of species (*curse of dimensionality*). For this reason, Monte Carlo approaches based on the stochastic simulation algorithm from [6] or related methods are often used. In an alternative line of research, numerical techniques have been applied to the CME in order to reduce the number of degrees of freedom, e.g. optimal state space truncation [1, 26, 27], spectral approximation [4], adaptive wavelet compression [16, 20], sparse grids [11], or tensor product approximation [2, 10, 18, 22, 21] among oth-

Tobias Jahnke
Karlsruhe Institute of Technology, Department of Mathematics, Kaiserstr. 93, D-76133 Karlsruhe, Germany, e-mail: tobias.jahnke@kit.edu

Vikram Sunkara
Karlsruhe Institute of Technology, Department of Mathematics, Kaiserstr. 93, D-76133 Karlsruhe, Germany, e-mail: vikram.sunkara@kit.edu

ers. But in spite of the progress achieved with these approaches, many biological systems are still out of reach of direct numerical approximation.

The size of the problem can be significantly reduced if only species with low populations are described by a probability distribution, whereas the abundant species are represented by (conditional) moments. This approach is motivated by the famous result in [23] which states, roughly speaking, that stochastic fluctuations in large populations are insignificant. In the last years, this has inspired the development of *hybrid models* where a low-dimensional CME is coupled to ordinary differential equations similar to the classical reaction-rate equation; cf. [5, 9, 11, 12, 13, 17, 25, 28].

In this article, we analyze the accuracy of a hybrid model called MRCE (*model reduction based on conditional expectations*). This approach has been proposed in [9, 17, 25], and it was demonstrated numerically that MRCE captures the critical bi-modal solution profiles which appear in certain applications. In [9, 17, 28], numerical techniques for MRCE were introduced, and an error bound for the modeling error was proven in [28]. In the present article, we make the additional assumption that the reaction system involves two scales, i.e. that the ratio between the small and large populations is proportional to a scaling parameter $0 < \varepsilon \ll 1$. For such two-scaled systems, we prove that the modeling error of the MRCE approximation is proportional to $\varepsilon$. The proof blends ideas and techniques from [19] and [28].

## 2 The Chemical Master Equation of two-scale reaction systems

We consider a partitioned reaction system with two groups of species denoted by $\mathsf{S}_1,\dots,\mathsf{S}_d$ and $\mathsf{S}_{d+1},\dots,\mathsf{S}_{d+D}$, respectively, with $d,D \in \mathbb{N}$. Let $X(t) \in \mathbb{N}_0^d$ be the vector whose entries $X_1(t),\dots X_d(t)$ indicate how many copies of each of the species $\mathsf{S}_1,\dots,\mathsf{S}_d$ exist at time $t \in [0,t_{\text{end}}]$, and let $Y(t) = (Y_1(t),\dots,Y_D(t))$ contain the copy numbers of $\mathsf{S}_{d+1},\dots,\mathsf{S}_{d+D}$. The species interact via $r \in \mathbb{N}$ reaction channels , $\mathsf{R}_1,\dots,\mathsf{R}_r$, each of which is represented by a scheme

$$\mathsf{R}_j : \sum_{k=1}^{d} a_{jk}\mathsf{S}_k + \sum_{k=1}^{D} b_{jk}\mathsf{S}_{d+k} \xrightarrow{c_j} \sum_{k=1}^{d} \widehat{a}_{jk}\mathsf{S}_k + \sum_{k=1}^{D} \widehat{b}_{jk}\mathsf{S}_{d+k}, \tag{1}$$

with $a_{jk},\widehat{a}_{jk},b_{jk},\widehat{b}_{jk} \in \mathbb{N}_0$ and $c_j > 0$. If the $j$-th reaction channel fires, then the population numbers jump from the current state $(X(t),Y(t)) = (n,m) \in \mathbb{N}_0^d \times \mathbb{N}_0^D$ to the new state $(n,m) + (\nu_j,\mu_j)$, where $(\nu_j,\mu_j) \in \mathbb{Z}^{d+D}$ is the stoichiometric vector associated to $\mathsf{R}_j$, i.e.

$$\nu_j = \left(\widehat{a}_{j1} - a_{j1},\dots,\widehat{a}_{jd} - a_{jd}\right)^T \in \mathbb{Z}^d$$
$$\mu_j = \left(\widehat{b}_{j1} - b_{j1},\dots,\widehat{b}_{jd} - b_{jd}\right)^T \in \mathbb{Z}^D.$$

In stochastic reaction kinetics, the function $t \mapsto (X(t),Y(t))$ is a realization of a Markov jump process; cf. [6, 14]. According to [6] the transition rates of this process depend on the propensity functions of the reaction channels. We assume that the propensity function of $\mathsf{R}_j$ has the form $\alpha_j(n)\beta_j(m)$ with

$$\alpha_j(n) = c_j \prod_{k=1}^{d} \binom{n_k}{a_{jk}}, \qquad \beta_j(m) = \varepsilon^{\gamma(j)-1}\varepsilon^{|b_j|} \prod_{k=1}^{D} \binom{m_k}{b_{jk}}, \tag{2}$$

where $|b_j| = \sum_{i=1}^{d} b_{ji}$, and where $0 < \varepsilon \ll 1$ is a scaling parameter discussed below. The value of $\gamma$ depends on whether or not the population numbers of the first group of species change when $\mathsf{R}_j$ fires. To be more precise, we partition the index set $\{1,\dots,r\}$ into

$$J_0 = \left\{ j \in \{1,\dots,r\} : \nu_j = (0,\dots,0)^T \right\}, \qquad J_1 = \{1,\dots,r\} \setminus J_0$$

and let $\gamma$ be the indicator function

$$\gamma(j) = \begin{cases} 0 \text{ if } j \in J_0, \\ 1 \text{ if } j \in J_1. \end{cases} \tag{3}$$

The reason for this particular scaling is the following: if $(X(t),Y(t)) = (n,m) \in \mathbb{N}_0^d \times \mathbb{N}_0^D$ with $n \in O(1)$ and $m \in O\left(\varepsilon^{-1}\right)$, then $\alpha_j(n)\beta_j(m) = O(c_j\varepsilon^{\gamma(j)-1})$ for all $j = 1,\dots,r$. Hence, the population numbers of $\mathsf{S}_{d+1},\dots,\mathsf{S}_{d+D}$ may change with a rate of $O\left(\varepsilon^{-1}\right)$, whereas the populations of $\mathsf{S}_1,\dots,\mathsf{S}_d$ only change with a rate of $O(1)$, provided that $c_j = O(1)$ for all $j$. For initial data $X(0) = O(1)$ and $Y(0) = O\left(\varepsilon^{-1}\right)$, one can thus expect that $\mathbb{E}(X(t)) = O(1)$ and $\mathbb{E}(Y(t)) = O\left(\varepsilon^{-1}\right)$ on bounded time intervals. Hence, $\varepsilon$ is roughly speaking the ratio between the small and the large population numbers of the two groups $\mathsf{S}_1,\dots,\mathsf{S}_d$ and $\mathsf{S}_{d+1},\dots,\mathsf{S}_{d+D}$, respectively. This scaling was extensively motivated and illustrated in [19], and a very similar scaling was considered in [25]. For $d = 0$ and $\alpha_j(n) = c_j$, our scaling coincides with the thermodynamic limit which has been analyzed in [23]. For $\varepsilon = 1$, there is no qualitative difference between the two groups of species, such that our setting corresponds to the situation in [6] where no partition nor scaling is considered.

Let $p(t,n,m)$ be the probability that at time $t \geq 0$ the process is in state $(n,m) \in \mathbb{N}_0^d \times \mathbb{N}_0^D$, i.e. the probability that there are $n_k$ copies of $\mathsf{S}_k$ for $k = 1,\dots,d$, and $m_l$ copies of $\mathsf{S}_{d+l}$ for $l = 1,\dots,D$. It is well-known (see [6, 7]) that the probability distribution $p$ evolves according to the Chemical Master Equation (CME)

$$\begin{aligned} \partial_t p(t,n,m) = \sum_{j=1}^{r} \Big( & \alpha_j(n-\nu_j)\beta_j(m-\mu_j)p(t,n-\nu_j,m-\mu_j) \\ & - \alpha_j(n)\beta_j(m)p(t,n,m) \Big) \qquad \forall (n,m) \in \mathbb{N}_0^d \times \mathbb{N}_0^D \end{aligned} \tag{4}$$

$$p(0,n,m) = p_0(n,m) \tag{5}$$

with the convention that $p(t,n-\nu_j,m-\mu_j)=0$ if $n-\nu_j \notin \mathbb{N}_0^d$ or $m-\mu_j \notin \mathbb{N}_0^D$. For the sake of a more compact notation we define the shift operators $S_j^1$ and $S_j^2$ by

$$S_j^1 u(n,m) = \begin{cases} u(n-\nu_j,m) & \text{if } n-\nu_j \in \mathbb{N}_0^d \\ 0 & \text{else} \end{cases}$$
$$S_j^2 u(n,m) = \begin{cases} u(n,m-\mu_j) & \text{if } m-\mu_j \in \mathbb{N}_0^D \\ 0 & \text{else} \end{cases}$$

for $u : \mathbb{N}_0^d \times \mathbb{N}_0^D \longrightarrow \mathbb{R}$. The two shift operators commute, i.e.

$$S_j^1 S_j^2 u(n,m) = S_j^2 S_j^1 u(n,m) = \begin{cases} u(n-\nu_j,m-\mu_j) & \text{if } n-\nu_j \in \mathbb{N}_0^d, m-\mu_j \in \mathbb{N}_0^D \\ 0 & \text{else.} \end{cases}$$

Products of functions are to be understood entry-wise, and applying a shift operator to a product $u(n,m)v(n,m)$ is to be understood in the sense that

$$\left(S_j^1 uv\right)(n,m) = \left(S_j^1(uv)\right)(n,m) = u(n-\nu_j,m)v(n-\nu_j,m) = \left(S_j^1 u\right)\left(S_j^1 v\right)(n,m).$$

With these operators, the CME (4) can be reformulated as

$$\partial_t p = \sum_{j=1}^{r} \left(S_j^1 S_j^2 - I\right)(\alpha_j \beta_j p). \tag{6}$$

The chemical master equation (6) is considered on the space

$$\ell^1 = \Big\{ u : \mathbb{N}_0^d \times \mathbb{N}_0^D \longrightarrow \mathbb{R} \;:\; \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} |u(n,m)| < \infty \Big\}.$$

of absolutely summable functions on $\mathbb{N}_0^d \times \mathbb{N}_0^D$. This is a straightforward extension of the standard $\ell^1$-space. For vector-valued functions $u=(u_1,\ldots,u_N) : \mathbb{N}_0^d \times \mathbb{N}_0^D \longrightarrow \mathbb{R}^N$ with some $N>1$, $u \in \ell^1$ means that $u_j \in \ell^1$ for all $j=1,\ldots,N$. The space $\ell^1$ is endowed with the norm

$$\|u\|_{\ell^1} = \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} |u(n,m)|$$

where $|\cdot| = |\cdot|_1$ is the 1-norm on $\mathbb{R}^N$. We set $\mathscr{X}^0 = \ell^1$ and define the spaces $\mathscr{X}^i$ via the recursion

$$\mathscr{X}^{i+1} = \Big\{ u \in \mathscr{X}^i \mid (n,m) \mapsto m_k u(n,m) \in \mathscr{X}^i \text{ for all } k \in \{1,\ldots,D\} \Big\}.$$

If $p(t,\cdot,\cdot) \in \ell^1$ is the solution of the CME (6), then $p_1(t,n) = \sum_m p(t,n,m)$ is the marginal distribution of $p(t,\cdot,\cdot)$, and if $p_1(t,n) \neq 0$, then

$$p_2(t,m \mid n) = \frac{p(t,n,m)}{p_1(t,n)} \tag{7}$$

is the conditional probability that at time $t$ there are $m_j$ particles of $\mathsf{S}_j$ given there are $n_i$ particles of $\mathsf{S}_i$ ($i \in \{1,\dots,d\}, j \in \{d+1,\dots,d+D\}$). If $p(t,\cdot,\cdot) \in \mathscr{X}^2$, then the conditional central moments

$$\begin{aligned} \xi(t,n) &= \sum_{m\in\mathbb{N}_0^D} m p_2(t,m \mid n) \\ C_\xi(t,n) &= \sum_{m\in\mathbb{N}_0^D} \big(m-\xi(t,n)\big)\big(m-\xi(t,n)\big)^T p_2(t,m \mid n) \end{aligned} \tag{8}$$

exist provided that $p_1(t,n) \neq 0$.

## 3 Model reduction based on conditional expectations

Solving the CME (4) or (6) numerically is a considerable challenge. First, the infinite state space $\mathbb{N}_0^d \times \mathbb{N}_0^D$ has to be truncated; this causes an error which has been analyzed in [26]. The truncated state space is finite, but still $(d+D)$-dimensional, and the total number of states is usually so large that standard numerical schemes cannot be applied.

On the other hand, the solution of the CME often provides more information than actually needed to understand the biological process. In many applications, one is mainly interested in the question how the stochastic behavior of $\mathsf{S}_1,\dots,\mathsf{S}_d$ affects the dynamics of $\mathsf{S}_{d+1},\dots,\mathsf{S}_{d+D}$. If the population numbers of $\mathsf{S}_{d+1},\dots,\mathsf{S}_{d+D}$ are sufficiently large, then stochastic fluctuations within their populations can be neglected according to [23]. In this case, it is sufficient to compute the *marginal distribution* $p_1(t,n)$ of the species $\mathsf{S}_1,\dots,\mathsf{S}_d$ along with conditional moments which measure the abundance of $\mathsf{S}_{d+1},\dots,\mathsf{S}_{d+D}$. This has motivated the construction of *hybrid models*: Instead of trying to solve the high-dimensional CME and then extracting the relevant information from the solution $p(t,n,m)$, one derives a reduced set of equations, namely a low-dimensional CME for the marginal distribution coupled with other ODEs; cf. [5, 9, 11, 12, 13, 17, 25, 28]. Hybrid models have the advantage that the huge number of unknowns is significantly reduced, which makes the problem computationally feasible. The price to pay is that hybrid models involve structurally more complicated differential equations than the (linear) CME, and that such a model reduction causes a modeling error in addition to the numerical error. The following hybrid model has been derived in [17]:

$$\partial_t w = \sum_{j\in J_1} (S_j^1 - I)\Big(\alpha_j \beta_j(\phi) w\Big) =: A(\phi) w \tag{9}$$

$$\begin{aligned} \partial_t(\phi w) &= \sum_{j\in J_1} (S_j^1 - I)\Big(\alpha_j \beta_j(\phi)\phi w\Big) + \sum_{j=1}^{r} \mu_j S_j^1 \Big(\alpha_j \beta_j(\phi) w\Big) \\ &=: F(\phi,w) + G(\phi,w) \end{aligned} \tag{10}$$

For fixed $\phi$, (9) is again a CME, but on the lower-dimensional state space $\mathbb{N}_0^d$. The function $w(t,n)$ approximates the marginal distribution $p_1(t,n)$ of the full CME solution, whereas $\phi(t,n)$ approximates the conditional expectations $\xi(t,n)$ defined in (8). This is why this model was called *model reduction based on conditional expectations (MRCE)* in [17]. It was demonstrated by numerical examples that MRCE captures certain bimodal solution profiles correctly, in contrast to simpler hybrid models proposed in the literature. Since $w$ and $\phi$ do not depend on $m$ any more, the $(d+D)$-dimensional state space of the CME is replaced by a $d$-dimensional state space, which reduces the computational costs considerably. Similar approaches have been proposed in [8, 9, 13, 15, 24, 25] for the CME and related differential equations.

Approximating the conditional expectations $\xi(t,n)$ has the drawback that (7) and hence (8) cannot be properly defined if $p_1(t,n)=0$. The same applies to the approximations $w(t,n)\approx p_1(t,n)$ and $\phi(t,n)\approx\xi(t,n)$. The hybrid model (9)-(10) is formulated in terms of $w$ and $\phi w$, but in order to evaluate the term $\beta_j(\phi)$ on the right-hand side, we have to divide $\phi(t,n)w(t,n)$ by $w(t,n)$. This is only possible if $w(t,n)>0$, and for $w(t,n)\approx 0$ such a division causes numerical instability. Different strategies to cope with this problem have been proposed in [9, 13, 17, 25, 28]. Since the main goal of the present article is an analysis of the accuracy of MRCE, we will avoid such technical problems by the following assumption:

**Assumption 1** *We assume that the CME (6) with initial condition (5) has a unique classical solution $p(t,\cdot,\cdot)\in\ell^1$ with strictly positive marginal distribution $p_1(t,\cdot)$, i.e. $p_1(t,n)>0$ for all $t\in[0,t_{end}]$ and all $n\in\mathbb{N}_0^d$. This implies that $p_2$, $\xi$ and $C_\xi$ are well-defined. Moreover, we assume that the hybrid model* (9)-(10) *with initial data*

$$w(0,n)=p_1(0,n) \qquad \text{and} \qquad \phi(0,n)=\xi(0,n) \tag{11}$$

*has a unique solution, and that $w(t,\cdot)\in\ell^1$ is strictly positive for all $t\in[0,t_{end}]$.*

This assumption seems to be a strong simplification because in typical applications it can be observed that for every threshold parameter $\delta\in(0,1)$, there are only finitely many states with $p(t,n,m)\geq\delta$. Roughly speaking, this means that $p(t,n,m)\approx 0$ for "most of" the states. However, if $p(t_\star,n_\star,m_\star)=0$ for some $t_\star>0$, then the state $(n_\star,m_\star)\in\mathbb{N}_0^d\times\mathbb{N}_0^D$ cannot be reached from neither of the states which had nonzero probability at time $t=0$. As a consequence, one could simply exclude $(n_\star,m_\star)$ from the state space to avoid the problem, and in this sense, Assumption 1 is not a severe restriction. Since numerical methods for solving (9)-(10) are not discussed in this article, numerical instabilities are not an issue here.

## 4 Error analysis for the hybrid model

Since $\beta_j(m)$ defined in (2) depends on the scaling parameter $\varepsilon$ and since $\beta_j(m)$ appears both in the CME (6) and in the hybrid model (9)-(10), the functions

$p(t,n,m)$, $p_1(t,n)$, $\xi(t,n)$, $w(t,n)$, and $\phi(t,n)$ all depend[1] on $\varepsilon$, too. In this section we prove that the modeling error of MRCE is bounded by $C\varepsilon$ (see Theorem 1 below). Throughout the article, $C$ denotes a generic constant which may have different values at different occurrences. The proof combines the arguments from [17, 19] with the analysis from [28] where systems with no scaling have been investigated. Our error analysis is based on the following assumptions.

**Assumption 2** *For every $j \in \{1,\dots,r\}$ we assume that $|b_j| \le 2$.*

This is a natural assumption, because the probability of a trimolecular reaction is negligible according to [7, page 418].

**Assumption 3** *We assume that the solution of the CME (6) satisfies $p(t,\cdot,\cdot) \in \mathscr{X}^3$ for $t \in [0,t_{end}]$ and that*

$$(n,m) \mapsto \alpha_j(n)p(t,n,m) \in \mathscr{X}^3 \qquad \text{for all } j \in \{1,\dots,r\}.$$

**Assumption 4** *We assume that*

$$\sup_{t\in[0,t_{end}]}\sup_{n\in\mathbb{N}_0^d}|\xi(t,n)| \le \frac{C}{\varepsilon}, \qquad \sup_{t\in[0,t_{end}]}\sup_{n\in\mathbb{N}_0^d}|C_\xi(t,n)| \le \frac{C}{\varepsilon},$$

$$\sup_{t\in[0,t_{end}]}\sup_{n\in\mathbb{N}_0^d}|\phi(t,n)| \le \frac{C}{\varepsilon}$$

*with a constant which does not depend on $\varepsilon$. Moreover, we assume that all third central moments of $p_2(t,\cdot \mid n)$ are bounded by $C\varepsilon^{-2}$ with a constant which does not depend on $t \in [0,t_{end}]$, $\varepsilon$, and $n \in \mathbb{N}_0^d$.*

**Assumption 5** *Suppose that there is a constant $C > 0$ such that for all $t \in [0,t_{end}]$ and $j \in \{1,\dots,r\}$ the bound*

$$\max_{j=1,\dots,r}\left\|\alpha_j(\cdot)u(t,\cdot)\right\|_{\ell^1} \le C\|u(t,\cdot)\|_{\ell^1}$$

*holds for each of the following functions:*

$$u = p_1, \qquad u = \beta_j(\xi)p_1 - \beta_j(\phi)w, \qquad u = \beta_j(\xi)\xi p_1 - \beta_j(\phi)\phi w.$$

Note that Assumption 4 implies $u \in \ell^1$ in each case.

The following error bound for the modeling error of MRCE is the main result of this article.

**Theorem 1.** *Under the assumptions 1, 2, 3, 4, and 5, there is a constant $C_b > 0$ such that the approximation error of MRCE is bounded by*

[1] We do not make this dependency explicit in the notation in order to keep the equations as simple as possible.

$$\sup_{t\in[0,t_{end}]} \|p_1(t,\cdot) - w(t,\cdot)\|_{\ell^1} \leq C_b\varepsilon \tag{12}$$

$$\sup_{t\in[0,t_{end}]} \|\xi(t,\cdot)p_1(t,\cdot) - \phi(t,\cdot)w(t,\cdot)\|_{\ell^1} \leq C_b. \tag{13}$$

*If in addition*

$$|b_j| \leq 1 \quad \textit{for all } j \in J_0, \qquad |b_j| = 0 \quad \textit{for all } j \in J_1, \tag{14}$$

*then MRCE is even exact, i.e. we can choose $C_b = 0$ in* (12) *and* (13)*.*

According to (13) the error of the approximation $\xi p_1 \approx \phi w$ remains bounded, but does not decrease when $\varepsilon \to 0$. This is not obvious, because Assumption 4 implies that $\|\xi(t,\cdot)p_1(t,\cdot)\|_{\ell^1} = O\left(\varepsilon^{-1}\right)$ and $\|\phi(t,\cdot)w(t,\cdot)\|_{\ell^1} = O\left(\varepsilon^{-1}\right)$. Multiplying both sides of (13) by $\varepsilon$ shows that the *relative error* converges linearly in $\varepsilon$.

**Proof.** It will be shown below in Lemma 2 and Lemma 3 that

$$\begin{aligned} &\|p_1(t,\cdot) - w_1(t,\cdot)\|_{\ell^1} + \varepsilon\,\|\xi p_1(t,\cdot) - \phi w(t,\cdot)\|_{\ell^1} \\ &\leq C_b\varepsilon + C\int_0^t \varepsilon\,\|(\xi p_1 - \phi w)(s,\cdot)\|_{\ell^1}\,ds + C\int_0^t \|p_1(s,\cdot) - w(s,\cdot)\|_{\ell^1}\,ds. \end{aligned} \tag{15}$$

for all $t \in [0,t_{\text{end}}]$ with constants $C_b$ and $C$ which do not depend on $t$ or $\varepsilon$. Hence, the Gronwall lemma yields

$$\|p_1(t,\cdot) - w_1(t,\cdot)\|_{\ell^1} + \varepsilon\,\|\xi p_1(t,\cdot) - \phi w(t,\cdot)\|_{\ell^1} \leq C_b\varepsilon$$

which proves (12) and (13). Moreover, it will be shown that we can choose $C_b = 0$ in the special case (14). □

The remainder of this article is devoted to the proof of the Gronwall inequality (15). As a preparatory step, we prove the following lemma:

**Lemma 1.** *Let $y : \mathbb{N}_0^d \longrightarrow \mathbb{R}^d$, $z : \mathbb{N}_0^d \longrightarrow \mathbb{R}^d$ with*

$$\max_{n\in\mathbb{N}_0^d} |y(n)| \leq C/\varepsilon, \qquad \max_{n\in\mathbb{N}_0^d} |z(n)| \leq C/\varepsilon, \tag{16}$$

*and let $u \in \ell^1$ and $v \in \ell^1$. Then for every $j \in \{1,\ldots,r\}$, there is a constant $C > 0$ such that*

$$\left\|\beta_j(y)u - \beta_j(z)v\right\|_{\ell^1} \leq C\varepsilon^{\gamma(j)}\left(\|yu - zv\|_{\ell^1} + \varepsilon^{-1}\|u - v\|_{\ell^1}\right)$$

*with $\gamma(j)$ defined in (3). Note that the assumption (16) implies that $yu - zv \in \ell^1$.*

A similar lemma has been shown in [19, Lemma 4].

**Proof.** For $|\lambda_j| = 0$ the assertion is obvious, because in this case $\beta_j(y) = \varepsilon^{\gamma(j)-1}$ is

constant. If $|\lambda_j| = 1$, then there is a $k \in \{1,\dots,d\}$ such that $\beta_j(y) = \varepsilon^{\gamma(j)} y_k$, and the assertion follows. If $|\lambda_j| = 2$, then the propensity $\beta_j(y)$ takes the form

$$\beta_j(y) = \hat{c}_j \varepsilon^{\gamma(j)+1} y_k y_l \qquad \text{with } \hat{c}_j = \begin{cases} c_j & \text{if } k \neq l \\ \frac{1}{2} c_j & \text{if } k = l \end{cases}$$

for some $k,l \in \{1,\dots,d\}$. Thus, we have to bound the difference

$$\begin{aligned} \beta_j(y)u - \beta_j(z)v &= \hat{c}_j \varepsilon^{\gamma(j)+1} \left(y_k y_l u - z_k z_l v\right) \\ &= \hat{c}_j \varepsilon^{\gamma(j)+1} \Big( y_k \left(y_l u - z_l v\right) + y_k z_l \left(v - u\right) + z_l \left(y_k u - z_k v\right) \Big). \end{aligned}$$

Since (16) implies that $|\varepsilon y_k(n)| \leq C$ and $|\varepsilon z_l(n)| \leq C$, it follows that

$$\left\| \beta_j(y)u - \beta_j(z)v \right\|_{\ell^1} \leq C\varepsilon^{\gamma(j)} \|yu - zv\|_{\ell^1} + C\varepsilon^{\gamma(j)-1} \|u - v\|_{\ell^1}$$

which proves the assertion. □

**Lemma 2.** *Under the assumptions of Theorem 1 there are constants $C_b \geq 0$ and $C > 0$ such that*

$$\begin{aligned} \|p_1(t,\cdot) - w_1(t,\cdot)\|_{\ell^1} \leq C_b\varepsilon &+ C \int_0^t \varepsilon \left\| \left(\xi p_1 - \phi w\right)(s,\cdot) \right\|_{\ell^1} \, ds \\ &+ C \int_0^t \|p_1(s,\cdot) - w(s,\cdot)\|_{\ell^1} \, ds. \end{aligned}$$

*for all $t \in [0, t_{end}]$. The constants $C_b$ and $C$ do not depend on $t$ or $\varepsilon$, and we can choose $C_b = 0$ in the special case* (14).

**Proof.** From the definition of the marginal distribution $p_1$ it follows that

$$\partial_t p_1 = \sum_{j \in J_0} \sum_{m \in \mathbb{N}_0^D} (S_j^1 S_j^2 - I)\alpha_j \beta_j p + \sum_{j \in J_1} \sum_{m \in \mathbb{N}_0^D} (S_j^1 S_j^2 - I)\alpha_j \beta_j p.$$

The first sum vanishes, because $S_j^1 = I$ for $j \in J_0$, and

$$\sum_{m \in \mathbb{N}_0^D} (S_j^2 - I)\alpha_j \beta_j p = 0 \tag{17}$$

by Lemma 2 in [19]. Since $S_j^1 S_j^2 - I = S_j^1 (S_j^2 - I) + (S_j^1 - I)$ and since (17) implies

$$\sum_{j \in J_1} \sum_{m \in \mathbb{N}_0^D} S_j^1 (S_j^2 - I)\alpha_j \beta_j p = 0,$$

we obtain

$$\partial_t p_1 \;=\; \sum_{j\in J_1}\sum_{m\in\mathbb{N}_0^D} (S_j^1 - I)\alpha_j\beta_j p \;=\; \sum_{j\in J_1}(S_j^1 - I)\alpha_j \sum_{m\in\mathbb{N}_0^D}\beta_j p_2 p_1. \tag{18}$$

Following the ideas of [3] we use the Taylor expansion

$$\beta_j(m) = \beta_j(\xi) + \nabla\beta_j(\xi)^T(m-\xi) + \frac{1}{2}(m-\xi)^T\left(\nabla^2\beta_j\right)(m-\xi) \tag{19}$$

where $\xi = \xi(t,n)$. Since $\beta_j$ is at most quadratic by Assumption 2, all higher-order terms vanish. This yields[2]

$$\sum_{m\in\mathbb{N}_0^D}\beta_j(m)p_2(t,m|n) = \beta_j(\xi) + R_j(t,n) \tag{20}$$

$$\text{with}\;\; R_j(t,n) = \operatorname{trace}\left(C_\xi(t,n)\nabla^2\beta_j\right)$$

because $\sum_{m\in\mathbb{N}_0^D} p_2(t,m|n) = 1$ and $\sum_{m\in\mathbb{N}_0^D}(m-\xi(t,n))p_2(t,m|n) = 0$. Substituting this into (18) gives

$$\partial_t p_1 \;=\; \sum_{j\in J_1}(S_j^1 - I)\alpha_j\beta_j(\xi)p_1 + \mathfrak{R} = A(\xi)p_1 + \mathfrak{R} \tag{21}$$

with a rest term $\mathfrak{R} = \mathfrak{R}(t,n)$ given by

$$\mathfrak{R} = \sum_{j\in J_1}(S_j^1 - I)\alpha_j R_j p_1.$$

Comparing (21) with (9) yields

$$\partial_t p_1 - \partial_t w = A(\xi)p_1 - A(\phi)w + \mathfrak{R}$$

and since $p_1(0,\cdot) = w(0,\cdot)$ according to (11), we obtain

$$\|p_1(t,\cdot) - w(t,\cdot)\|_{\ell^1} \le \int_0^t \left\|A\big(\xi(s,\cdot)\big)p_1(s,\cdot) - A\big(\phi(s,\cdot)\big)w(s,\cdot)\right\|_{\ell^1} ds \tag{22a}$$

$$+ \int_0^t \left\|\mathfrak{R}(s,\cdot)\right\|_{\ell^1} ds. \tag{22b}$$

Our next goal is to derive a bound for the second term (22b). According to Assumption 4 we have

$$\sup_{t\in[0,t_{\text{end}}]}\sup_{n\in\mathbb{N}_0^d}|C_\xi(t,n)| \le \frac{C}{\varepsilon},$$

whereas (2) yields

[2] The remainder term $R_j$ is not to be mixed up with the reaction channel $\mathsf{R}_j$ in (1).

$$\nabla^2\beta_j = \begin{cases} 0 & \text{if } |b_j| \le 1 \\ \varepsilon^{\gamma(j)+1} & \text{if } |b_j| = 2. \end{cases}$$

By Assumption 2, no other cases have to be considered. Hence, it follows that

$$\sup_{s\in[0,t_{\text{end}}]}\sup_{n\in\mathbb{N}_0^d}|R_j(s,n)| = \sup_{s\in[0,t_{\text{end}}]}\sup_{n\in\mathbb{N}_0^d}\left|\operatorname{trace}\left(C_\xi(s,n)\nabla^2\beta_j\right)\right| \le C\varepsilon^{\gamma(j)}, \tag{23}$$

and Assumption 5 and the fact that $\gamma(j) = 1$ for all $j \in J_1$ yield the estimate

$$\int_0^t \left\|\mathfrak{R}(s,\cdot)\right\|_{\ell^1} ds \le Ct\varepsilon \sup_{s\in[0,t]}\sup_{j\in J_1}\|\alpha_j(s,\cdot)p_1(s,\cdot)\|_{\ell^1} \le C_b t_{\text{end}}\varepsilon. \tag{24}$$

If $|b_j| \in \{0,1\}$ for all $j = 1,\dots,r$, then $\nabla^2\beta_j = 0$ and hence $\left\|\mathfrak{R}(s,\cdot)\right\|_{\ell^1} = 0$ such that one can choose $C_b = 0$ in the special case (14). The first error term (22a) can be bounded by

$$\begin{aligned}
&\left\|A(\xi)p_1 - A(\phi)w\right\|_{\ell^1} \\
&= \Big\| \sum_{j\in J_1}(S_j^1 - I)\Big(\beta_j(\xi)\alpha_j p_1\Big) - \sum_{j\in J_1}(S_j^1 - I)\Big(\beta_j(\phi)\alpha_j w\Big)\Big\|_{\ell^1} \\
&\le C\max_{j\in J_1}\left\|\beta_j(\xi)\alpha_j p_1 - \beta_j(\phi)\alpha_j w\right\|_{\ell^1} \\
&\le C\max_{j\in J_1}\left\|\beta_j(\xi)p_1 - \beta_j(\phi)w\right\|_{\ell^1}
\end{aligned} \tag{25}$$

due to Assumption 5. Applying Lemma 1 now yields

$$\left\|\beta_j(\xi)p_1 - \beta_j(\phi)w\right\|_{\ell^1} \le C\varepsilon^{\gamma(j)}\left(\left\|\xi p_1 - \phi w\right\|_{\ell^1} + \varepsilon^{-1}\left\|p_1 - w\right\|_{\ell^1}\right),$$

and since the maximum in (25) is only taken over $J_1$, it follows that

$$\left\|A(\xi)p_1 - A(\phi)w\right\|_{\ell^1} \le C\left(\varepsilon\left\|\xi p_1 - \phi w\right\|_{\ell^1} + \left\|p_1 - w\right\|_{\ell^1}\right). \tag{26}$$

Substituting (24) and (26) into (22a) and (22b) yields the assertion. □

**Lemma 3.** *Under the assumptions of Theorem 1 there are constants $C_b \ge 0$ and $C > 0$ such that*

$$\begin{aligned}
\|(\xi p_1 - \phi w)(t,\cdot)\|_{\ell^1} \le C_b &+ C\int_0^t \|(\xi p_1 - \phi w)(s,\cdot)\|_{\ell^1}\, ds \\
&+ \frac{C}{\varepsilon}\int_0^t \|p_1(s,\cdot) - w(s,\cdot)\|_{\ell^1}\, ds.
\end{aligned}$$

*for all $t \in [0,t_{end}]$. The constants $C_b$ and $C$ do not depend on $t$ or $\varepsilon$. If $|b_j| \in \{0,1\}$ for all $j = 1,\dots,r$, then we can choose $C_b = 0$.*

**Proof.** With similar arguments as in the proof of Lemma 2, it can be shown that

$$\begin{aligned}\partial_t\left(\xi p_1\right)(t,n) = &\sum_{j=1}^{r}\mu_j \sum_{m\in\mathbb{N}_0^D}\left(S_j^1\alpha_j\beta_j p\right)(t,n,m)\\ &+\sum_{j\in J_1}\sum_{m\in\mathbb{N}_0^D} m\left((S_j^1-1)\alpha_j\beta_j p\right)(t,n,m)\end{aligned} \tag{27}$$

(see step 1 in the proof of Lemma 6 in [19] for details). For the first term on the right-hand side, (20) yields

$$\begin{aligned}\sum_{m\in\mathbb{N}_0^D}\left(S_j^1\alpha_j\beta_j p\right)(t,n,m) &= \alpha_j(n-\nu_j)\left(\sum_{m\in\mathbb{N}_0^D}\beta_j(m)p_2(t,m\mid n-\nu_j)\right)p_1(t,n-\nu_j)\\ &= \alpha_j(n-\nu_j)\Big(\beta_j\big(\xi(t,n-\nu_j)\big)+R_j(t,n-\nu_j)\Big)p_1(t,n-\nu_j)\\ &= S_j^1\big(\alpha_j\big[\beta_j(\xi)+R_j\big]p_1\big)(t,n).\end{aligned}$$

Moreover, it follows from (19) that

$$\sum_{m\in\mathbb{N}_0^D} m\beta_j(m)p(t,n,m) = \Big(\beta_j(\xi)\xi+T_j(t,n)\Big)p_1(t,n)$$

with $\xi=\xi(t,n)$ and

$$\begin{aligned}T_j(t,n) =& C_\xi(t,n)\nabla\beta_j(\xi)+\frac{1}{2}\xi R_j(t,n)\\ &+\frac{1}{2}\sum_{m\in\mathbb{N}_0^D}(m-\xi)(m-\xi)^T(\nabla^2\beta_j)(m-\xi)p_2(t,m\mid n).\end{aligned}$$

Substituting into (27) yields

$$\begin{aligned}\partial_t\left(\xi p_1\right)(t,n) = &\sum_{j=1}^{r}\mu_j S_j^1\big(\alpha_j\beta_j(\xi)p_1\big)(t,n)\\ &+\sum_{j\in J_1}\left((S_j^1-1)\alpha_j\beta_j(\xi)\xi p_1\right)(t,n)+\mathfrak{R}(t,n)\end{aligned}$$

with defect

$$\mathfrak{R}(t,n) = \sum_{j=1}^{r}\mu_j S_j^1\big(R_j\alpha_j p_1\big)(t,n)+\sum_{j\in J_1}\left((S_j^1-1)T_j\alpha_j p_1\right)(t,n).$$

Comparing this with (10) shows that

$$\partial_t(\xi p_1)(t,n) = F(\xi,p_1)+G(\xi,p_1)+\mathfrak{R}.$$

We will now prove that $\|\mathfrak{R}(t,\cdot)\|_{\ell^1} \leq C_b$ with a constant $C_b \geq 0$ which does not depend on $\varepsilon$ nor on $t \in [0,t_{\text{end}}]$. In the special case (14), we have that $\nabla\beta_j(\xi) = 0$ for all $j \in J_1$ and $\nabla^2\beta_j(\xi) = 0$ for all $j = 1,\ldots,r$. This implies $R_j = 0$ for all $j$, $T_j = 0$ for all $j \in J_1$ and hence $\left\|\mathfrak{R}(s,\cdot)\right\|_{\ell^1} = 0$ such that one can choose $C_b = 0$. If (14) is not true, then according to (23) we know that $|R_j(s,n)| \leq C$ for all $j = 1,\ldots,r$, and with straightforward calculations and Assumption 5 we obtain the bound

$$\Big\| \sum_{j=1}^{r} \mu_j S_j^1 \big(R_j \alpha_j p_1\big)(t,\cdot)\Big\|_{\ell^1} \leq C \max_{j=1,\ldots,r} \|\alpha_j(\cdot)p_1(t,\cdot)\|_{\ell^1} \leq C.$$

Concerning the second term in $\mathfrak{R}$, Assumption 4 and (2) imply that

$$\sup_{t\in[0,t_{\text{end}}]} \sup_{n\in\mathbb{N}_0^d} |T_j(t,n)| \leq C\varepsilon^{\gamma(j)-1}.$$

The sum in the second term of $\mathfrak{R}$ is only taken over $j \in J_1$ such that $\gamma(j) = 1$. With Assumption 5, it thus follows that

$$\Big\| \sum_{j\in J_1} \sum_{m\in\mathbb{N}_0^D} \big((S_j^1 - 1)T_j\alpha_j p_1\big)(t,\cdot)\Big\|_{\ell^1} \leq C \max_{j=1,\ldots,r} \|\alpha_j(\cdot)p_1(t,\cdot)\|_{\ell^1} \leq C,$$

which proves $\|\mathfrak{R}(t,\cdot)\|_{\ell^1} \leq C_b$. Now the error $\xi p_1 - \phi w$ can be estimated by

$$\begin{aligned}
\|(\xi p_1)(t,\cdot) - (\phi w)(t,\cdot)\|_{\ell^1} &\leq \int_0^t \|\partial_t(\xi p_1)(s,\cdot) - \partial_t(\phi w)(s,\cdot)\|_{\ell^1}\, ds \\
&\leq \int_0^t \|F(\xi p_1)(s,\cdot) - F(\phi w)(s,\cdot)\|_{\ell^1}\, ds \\
&\quad + \int_0^t \|G(\xi p_1)(s,\cdot) - G(\phi w)(s,\cdot)\|_{\ell^1}\, ds + Ct.
\end{aligned} \tag{28}$$

It follows from Assumption 5 and Lemma 1 that

$$\begin{aligned}
&\|G(\xi,p_1)(s,\cdot) - G(\phi,w)(s,\cdot)\|_{\ell^1} \\
&\leq C \max_{j=1,\ldots,r} \big\|\alpha_j\big[\beta_j(\xi)p_1 - \beta_j(\phi)w\big](s,\cdot)\big\|_{\ell^1} \\
&\leq C \max_{j=1,\ldots,r} \big\|\big[\beta_j(\xi)p_1 - \beta_j(\phi)w\big](s,\cdot)\big\|_{\ell^1} \\
&\leq C\,\|\xi p_1(s,\cdot) - \phi w(s,\cdot)\|_{\ell^1} + \frac{C}{\varepsilon}\,\|p_1(s,\cdot) - w(s,\cdot)\|_{\ell^1}\,.
\end{aligned} \tag{29}$$

A corresponding bound has to be shown for $F(\xi p_1) - F(\phi w)$. Assumption 5 yields

$$\begin{aligned}
\|F(\xi p_1)(s,\cdot) - F(\phi w)(s,\cdot)\|_{\ell^1} &\leq C \max_{j\in J_1} \Big\|\alpha_j\beta_j(\xi)\xi p_1(s,\cdot) - \alpha_j\beta_j(\phi)\phi w(s,\cdot)\Big\|_{\ell^1} \\
&\leq C \max_{j\in J_1} \Big\|\beta_j(\xi)\xi p_1(s,\cdot) - \beta_j(\phi)\phi w(s,\cdot)\Big\|_{\ell^1}.
\end{aligned}$$

We decompose the error into three parts:

$$\begin{aligned}
&\|F(\xi p_1)(s,\cdot) - F(\phi w)(s,\cdot)\|_{\ell^1} \\
&\leq C \max_{j\in J_1} \Big( \Big\| \beta_j(\xi)[\xi p_1 - \phi w](s,\cdot) \Big\|_{\ell^1} + \Big\| \beta_j(\xi)\phi[w-p_1](s,\cdot) \Big\|_{\ell^1} \\
&\qquad + \Big\| \phi[\beta_j(\xi)p_1 - \beta_j(\phi)w](s,\cdot) \Big\|_{\ell^1} \Big).
\end{aligned}$$

Since (2) and Assumption 4 imply that for every $j \in J_1$

$$\begin{aligned}
\sup_{t\in[0,t_{\mathrm{end}}]} \sup_{n\in\mathbb{N}_0^d} \big|\beta_j\big(\xi(t,n)\big)\big| &\leq C\varepsilon^{\gamma(j)-1}\varepsilon^{|b_j|} \sup_{t\in[0,t_{\mathrm{end}}]} \sup_{n\in\mathbb{N}_0^d} \big|\big(\xi(t,n)\big)\big|^{|b_j|} \\
&\leq C\varepsilon^{\gamma(j)-1} = C,
\end{aligned}$$

and since $\sup_{n\in\mathbb{N}_0^d} |\phi(s,n)| \leq \frac{C}{\varepsilon}$ by Assumption 4, we obtain

$$\begin{aligned}
&\|F(\xi p_1)(s,\cdot) - F(\phi w)(s,\cdot)\|_{\ell^1} \\
&\leq C \Big\| [\xi p_1 - \phi w](s,\cdot) \Big\|_{\ell^1} \\
&\qquad + \frac{C}{\varepsilon} \Big( \big\| w(s,\cdot) - p_1(s,\cdot) \big\|_{\ell^1} + \max_{j\in J_1} \Big\| \big[\beta_j(\xi)p_1 - \beta_j(\phi)w\big](s,\cdot) \Big\|_{\ell^1} \Big).
\end{aligned}$$

Applying Lemma 1 to the last term yields

$$\max_{j\in J_1} \Big\| \big[\beta_j(\xi)p_1 - \beta_j(\phi)w\big](s,\cdot) \Big\|_{\ell^1} \leq C\varepsilon \| \big[\xi p_1 - \phi w\big] \|_{\ell^1} + C\|p_1 - w\|_{\ell^1}$$

because $\gamma(j) = 1$ for $j \in J_1$. Hence, we have shown the estimate

$$\begin{aligned}
\|F(\xi p_1)(s,\cdot) - F(\phi w)(s,\cdot)\|_{\ell^1} \leq & C \big\| \xi p_1(s,\cdot) - \phi w(s,\cdot) \big\|_{\ell^1} \\
& + \frac{C}{\varepsilon} \big\| w(s,\cdot) - p_1(s,\cdot) \big\|_{\ell^1}
\end{aligned} \tag{30}$$

Substituting (29) and (30) into (28) proves the assertion. □

## References


1. Burrage, K., Hegland, M., MacNamara, S., Sidje, R.B.: A Krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems. In: A.N.Langville, W.J.Stewart (eds.) Markov Anniversary Meeting: An international conference to celebrate the 150th anniversary of the birth of A.A. Markov, pp. 21 – 38. Boson Books (2006)
2. Dolgov, S.V., Khoromskij, B.N.: Tensor-product approach to global time-space-parametric discretization of chemical master equation. Tech. rep., Max-Planck-Institut für Mathematik in den Naturwissenschaften (2012)

3. Engblom, S.: Computing the moments of high dimensional solutions of the master equation. Appl. Math. Comput. **180**(2), 498–515 (2006)
4. Engblom, S.: Spectral approximation of solutions to the chemical master equation. J. Comput. Appl. Math. **229**(1), 208–221 (2009)
5. Ferm, L., Lötstedt, P., Hellander, A.: A hierarchy of approximations of the master equation scaled by a size parameter. J. Sci. Comput. **34**, 127–151 (2008)
6. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. **22**, 403–434 (1976)
7. Gillespie, D.T.: A rigorous derivation of the chemical master equation. Physica A **188**, 404–425 (1992)
8. Griebel, M., Jager, L.: The BGY3dM model for the approximation of solvent densities. J. Chem. Phys. **129**(17), 174,511–174,525 (2008)
9. Hasenauer, J., Wolf, V., Kazeroonian, A., Theis, F.: Method of conditional moments for the chemical master equation. Tech. rep., to appear in J. Math. Biol. (2013)
10. Hegland, M., Garcke, J.: On the numerical solution of the chemical master equation with sums of rank one tensors. In: W. McLean, A.J. Roberts (eds.) Proceedings of the 15th Biennial Computational Techniques and Applications Conference, CTAC-2010, pp. C628–C643 (2011)
11. Hegland, M., Hellander, A., Lötstedt, P.: Sparse grids and hybrid methods for the chemical master equation. BIT **48**, 265–284 (2008)
12. Hellander, A., Lötstedt, P.: Hybrid method for the chemical master equation. J. Comput. Phys. **227**, 100–122 (2007)
13. Henzinger, T., Mateescu, M., Mikeev, L., Wolf, V.: Hybrid numerical solution of the chemical master equation. In: P. Quaglia (ed.) Proceedings of the 8th International Conference on Computational Methods in Systems Biology (CMSB10), pp. 55–65. ACM (2010)
14. Higham, D.J.: Modeling and simulating chemical reactions. SIAM Rev. **50**(2), 347–368 (2008)
15. Iedema, P.D., Wulkow, M., Hoefsloot, H.C.J.: Modeling molecular weight and degree of branching distribution of low-density polyethylene. Macromolecules **33**, 7173–7184 (2000)
16. Jahnke, T.: An adaptive wavelet method for the chemical master equation. SIAM J. Sci. Comput. **31**(6), 4373–4394 (2010)
17. Jahnke, T.: On reduced models for the chemical master equation. SIAM Multiscale Model. Simul. **9**(4), 1646–1676 (2011)
18. Jahnke, T., Huisinga, W.: A dynamical low-rank approach to the chemical master equation. Bull. Math. Biol. **70**(8), 2283–2302 (2008)
19. Jahnke, T., Kreim, M.: Error bound for piecewise deterministic processes modeling stochastic reaction systems. SIAM Multiscale Model. Simul. **10**(4), 1119–1147 (2012)
20. Jahnke, T., Udrescu, T.: Solving chemical master equations by adaptive wavelet compression. J. Comput. Phys. **229**(16), 5724–5741 (2010)
21. Kazeev, V., Khammash, M., Nip, M., Schwab, C.: Direct solution of the chemical master equation using quantized tensor trains. Tech. rep., ETH Zurich (2013)
22. Kazeev, V., Schwab, C.: Tensor approximation of stationary distributions of chemical reaction networks. Tech. rep., ETH Zurich (2013)
23. Kurtz, T.G.: The relationship between stochastic and deterministic models of chemical reactions. J. Chem. Phys. **57**, 2976–2978 (1973)
24. Lötstedt, P., Ferm, L.: Dimensional reduction of the Fokker-Planck equation for stochastic chemical reactions. Multiscale Model. Simul. **5**, 593–614 (2006)
25. Menz, S., Latorre, J.C., Schütte, C., Huisinga, W.: Hybrid stochastic-deterministic solution of the chemical master equation. SIAM Multiscale Model. Simul. **10**(4), 1232–1262 (2012)
26. Munsky, B., Khammash, M.: The finite state projection algorithm for the solution of the chemical master equation. J. Chem. Phys. **124**(4), 044,104 (2006)
27. Sunkara, V.: Analysis and numerics of the chemical master equation. Ph.D. thesis, Australian National University (2013)
28. Sunkara, V.: Finite state projection method for hybrid models. Tech. rep., Karlsruhe Institute of Technology (2013)