A dynamical low-rank approach to the chemical master equation

Tobias Jahnke and Wilhelm Huisinga

Tobias Jahnke

Universität Karlsruhe (TH)
Institut für Angewandte und
Numerische Mathematik
Karlsruhe, Germany

jahnke@math.uni-karlsruhe.de

Wilhelm Huisinga

Hamilton Institute, NUIM
Maynooth, Co. Kildare
Ireland

and

DFG Research Center MATHEON
Berlin, Germany

wilhelm.huisinga@nuim.ie

Corresponding author:

Tobias Jahnke
Universität Karlsruhe (TH)
Institut für Angewandte und Numerische Mathematik
Englerstr. 2
D-76128 Karlsruhe

phone: +49 (0)721 608 7982
Fax.: +49 (0)721 608 3197

e-mail: `jahnke@math.uni-karlsruhe.de`

The final manuscript will be submitted electronically in LaTeX.

A dynamical low-rank approach to the chemical master equation

Tobias Jahnke[1] and Wilhelm Huisinga[2]

Abstract

Stochastic reaction kinetics has increasingly been used to study cellular systems, with applications ranging from viral replication to gene regulatory networks and to signalling pathways. The underlying evolution equation, known as the chemical master equation (CME), can rarely be solved with traditional methods due to the huge number of degrees of freedom. We present a new approach to directly solve the CME by a dynamical low-rank approximation based on the Dirac-Frenkel-McLachlan variational principle. The new approach has the capability to substantially reduce the number of degrees of freedom, and to turn the CME into a computationally tractable problem. We illustrate the accuracy and efficiency of our methods in application to two examples of biological interest.

1   Introduction

It is widely recognized that the traditional reaction-rate approach to cellular reaction systems in terms of ordinary differential equation is inappropriate, if (i) some of the substances are present in a low number of molecules, or (ii) fluctuations play an important role in the evolution of the system [40, 2, 39]. This is often the case in the investigation of gene regulatory networks and signalling pathways, where both the discrete and the stochastic nature of such systems have to be taken into account [11, 35, 28, 29]. The stochastic description of chemical reaction kinetics is given by the chemical master equation (CME) which governs the evolution of the probability distribution describing the state of the reaction system (e.g., [16, 13]). Only rarely the CME can be solved analytically [22], so numerical methods have to be devised to approximate its solution.

Monte Carlo based approaches are typically employed to generate a statistically large ensemble of realizations of the stochastic process associated with the CME. The most prominent of these approaches is the stochastic simulation algorithm introduced by Gillespie [14, 15]. Since its computational effort scales with the number of reactions occurring, the stochastic simulation algorithm can be to expensive for highly reactive biological systems. As a remedy, approximate techniques have been suggested based on tau-leaping or partitioning of the system [17, 36, 7, 5, 18, 37, 34, 20, 1, 38, 6, 10]. Tau-leaping methods assume

---

[1]Universität Karlsruhe (TH), Institut für Angewandte und Numerische Mathematik, Englerstr. 2, D-76128 Karlsruhe, `jahnke@math.uni-karlsruhe.de`

[2]Hamilton Institute, NUIM, Maynooth, Co. Kildare, Ireland and DFG Research Center MATHEON, Berlin/Germany, `wilhelm.huisinga@nuim.ie`

that many reaction events will occur without significantly changing the reaction propensities, allowing to locally approximate the inhomogenous Poisson process by a simpler distribution [17, 36, 7, 5]. Systems-partitioning methods rely on the existence of multiple time scales, separating the system into slow and fast reactions. In [18, 37, 34] the fast reactions are solved analytically using some quasi-steady state assumption, resulting in a reduced stochastic system of slow reactions to be solved. Rather than explicitly computing the quasi steady states, a different line of approaches solves the fast reaction system numerically, either using multi-scale techniques, or approximating the fast dynamics by some Langevin or deterministic reaction rate equation [20, 1, 38, 6, 10]. In order to analyze the stochastic reaction system, all above Monte Carlo approaches require a statistically large number of realizations. For some systems, important biological events may be very rare, and extremely large number of realizations may be required to make the statistics sufficiently precise.

A different line of approach is to directly solve the CME as a linear differential equation. For many biological systems it is observed that the probability distribution is typically very small outside a bounded subdomain. Thus, it seems reasonable to truncate the state space to a finite, yet sufficiently large domain and approximate the exact solution by the solution of the resulting truncated CME. In [31], the finite state projection method was introduced that provides bounds on the approximation error as well as methods for systematically increasing the state space to meet a pre-defined accuracy. While truncation of the state space results in a finite dimensional linear system of ordinary differential equations (ODEs), the state space is typically still too huge for standard approaches to be applicable—demanding for efficient numerical techniques for its solution. Numerical approaches can roughly be grouped into two categories: Grid-based methods [33, 31, 4, 27] and Galerkin-based methods [9, 12, 8].

Grid-based approaches explicitly represent the probability distribution on the finite domain. In [33], the CME was solved directly via computation of the matrix-exponential, exploiting multiple time scales and singular perturbation theory. A Krylov-based approach to approximate the probability distribution in time has been exploited in [4] and [27], whereas a sparse grid approach to solve the truncated CME was advocated in [21]. Galerkin-based approaches represent the probability distribution in terms of some (adaptively) chosen finite dimensional ansatz space. In [9, 12], the solution is represented with respect to a time-dependent orthogonal basis of discrete Charlier polynomials adapted over time. This interesting method works well as long as the solution of the CME is sufficiently smooth and nearly unimodal. However, bimodality, as opposed to unimodality, and low regularity are often crucial properties of the CME solution, and in this situation, the representation with respect to Charlier polynomials can become inefficient. In [8] an adaptive Galerkin method is proposed that avoids these drawbacks by discretizing the CME in the Rothe context and replacing the global polynomial basis by a local one according to an adaptive decomposition of the state space.

In this paper we present a novel approach to directly solve the CME. It is based on a dynamical low-rank approximation (DLRA) which allows to re-

duce the number of degrees of freedom of the CME. The underlying idea is to project the CME to a manifold of suitably chosen basis functions and derive new equations of motion for the basis functions as well as the expansion coefficients from the Dirac-Frenkel-McLachlan variational principle. Similar methods have been used with great success in computational chemistry to approximately solve high-dimensional Schrödinger equations (see, e.g., [30, 3] and references therein) and have recently been investigated from the perspective of numerical analysis in [23]. In Sec. 2 we introduce the CME and explain the numerical challenges related to its solution in more detail. Then, based on the variational principle, the dynamical low-rank approximation to the CME is derived in Section 3. In Sec. 4, the new approach is applied to a genetic toggle switch and to a model for the life cycle of bacteriophage-λ. The final section concludes with a discussion and possible extensions of the presented approach.

2　The chemical master equation

We consider a well-stirred chemical reaction system of $d \in \mathbb{N}$ species $S_1, \ldots, S_d$ interacting through $N \in \mathbb{N}$ reaction channels. In the stochastic description the system is represented by a vector-valued random variable

$$X(t) = (X_1(t), \ldots, X_d(t))$$

on the discrete state space $\mathbb{N}^d$ where $\mathbb{N}$ is meant to include the number 0. The entry $X_j(t)$ describes the number of molecules of the j-th species at time t. It is well known that the probability distribution

$$P(t,x) = \mathbb{P}\Big(X_1(t) = x_1, \ldots, X_d(t) = x_n\Big), \qquad x = (x_1, \ldots, x_n) \in \mathbb{N}^d$$

of the random variable $X(t)$ is the solution of the chemical master equation (CME)

$$\partial_t P(t,x) = \sum_{m=1}^{N} \Big(\alpha_m(x - \nu_m)P(t, x - \nu_m) - \alpha_m(x)P(t,x) \Big) \qquad (1)$$

(cf. [14, 16, 13]). Here, $\nu_m \in \mathbb{Z}^d$ is the stoichiometric vector of the m-th reaction and $\alpha_m : \mathbb{N}^d \to [0, \infty)$ denotes the corresponding propensity function. The first term on the right-hand side has to be omitted if the argument $x - \nu_m$ contains negative entries. Therefore, we define

$$P(t,x) = 0 \quad \text{and} \quad \alpha_m(x) = 0 \quad \text{for all} \quad x \notin \mathbb{N}^d. \qquad (2)$$

Eq. (1) can be rewritten as

$$\partial_t P(t, \cdot) = \mathcal{A}P(t, \cdot) \qquad (3)$$

where $\mathcal{A}$ denotes the linear operator

$$\big(\mathcal{A}P(t, \cdot)\big)(x) = \sum_{m=1}^{N} \alpha_m(x - \nu_m)P(t, x - \nu_m) - \alpha_m(x)P(t,x). \qquad (4)$$

The CME can be regarded as an infinite system of ordinary differential equations (ODEs), or as a discrete partial differential equation with spatial differences instead of derivatives. In order to turn the CME into a finite problem, a finite state space projection (FSP) was suggested in [31]. This method aims at truncating the state space to a finite sub-system which still captures enough of the information and at the same time allows for a numerical solution. The truncation error is estimated as follows: Let us denote by

$$\Omega_\xi = \{x \in \mathbb{N}^d \; : \; 0 \le x_i \le \xi_i \text{ for } i = 1, \, \ldots \, , d\}, \qquad \xi = (\xi_1, \ldots, \xi_d)$$

the truncated state space, by $\mathcal{A}_\xi$ the restriction of $\mathcal{A}$ to Ω_ξ and by $P_\xi(t)$ the solution of the restricted CME $\partial_t P_\xi(t) = \mathcal{A}_\xi P_\xi(t)$ with initial condition $P_\xi(0)$, where $P_\xi(0)$ denotes the initial density restricted to the truncated state space. Define $P_\xi(t, \Omega_\xi) = \sum_{x \in \Omega_\xi} P_\xi(t, x)$. The FSP method states that if $P_\xi(t, \Omega_\xi) \ge 1 - \varepsilon$, then

$$P_\xi(t, x) \le P(t, x) \le P_\xi(t, x) + \varepsilon \qquad \text{for } x \in \Omega_\xi. \tag{5}$$

This estimate indicates how close the solution $P_\xi(t)$ on the truncated state space approximates the true solution. To make use of the FSP two problems have to be addressed: (i) Given $\varepsilon > 0$ and the final time t, how to find an appropriate truncated state space such that $P_\xi(t, \Omega_\xi) \ge 1 - \varepsilon$? (ii) How to efficiently compute the solution of the CME on the truncated state space which may still involve a huge number of states? In [31] and [33], the sub-problem (i) was tackled by exploiting the concept of reachability, or biological insight into the system. Regarding the sub-problem (ii), different approaches have been suggested, including the direct solution of the truncated system via the matrix-exponential [31], potentially in combination with exploitation of multiple time scales [33], a Krylov-based approach [4, 27] or a sparse grid approach [21]. In this article, we advocate a new approach to solve the truncated state space CME (sub-problem (ii)). It is based on a dynamical low-rank approximation in combination with the Dirac-Frenkel-McLachlan variational principle.

3 Dynamical low-rank approximations (DLRA)

3.1 Notation

In the sequel we assume that the state space has already been truncated and omit the subscripts ξ referring to the finite state projection. On the truncated state space the solution $P : \mathbb{R} \times \Omega_\xi \longrightarrow [0, 1]$ can be regarded either as a time-dependent function of a discrete variable or as a time-dependent, d-dimensional tensor because $P(t, \, \cdot \,)$ assigns a real number to every multi-index $x \in \Omega_\xi \subset \mathbb{N}^d$. Since both perspectives have certain advantages we will frequently switch between both of them.

For any function $F : \mathbb{R} \times \Omega_\xi \longrightarrow \mathbb{R}$ let $F(t)$ denote the function $F(t, \, \cdot \,)$:

$\Omega_\xi \longrightarrow [0,1]$ with fixed t. Moreover, we introduce the inner product

$$\langle F \ , \ G \rangle \quad = \quad \sum_{x \in \Omega_\xi} F(x)G(x) \tag{6}$$

for functions $F, G : \Omega_\xi \longrightarrow \mathbb{R}$. If F and G do not depend on all entries of x, then the sum is assumed to be taken over all common spatial directions. For example, if $F = F(x_1, x_2, x_4)$ and $G = G(x_2, x_3, x_4)$ then

$$\langle F \ , \ G \rangle \quad = \quad \sum_{x_2=0}^{\xi_2} \sum_{x_4=0}^{\xi_4} F(x_1, x_2, x_4)G(x_2, x_3, x_4).$$

Note that in this case the result is not a scalar but a function with respect to x_1 and x_3. For all functions $F : \Omega_\xi \longrightarrow \mathbb{R}$ the inner product (6) induces the norm

$$\|F\|_2 \quad = \quad \sqrt{\langle F \ , \ F \rangle}$$

which can be seen as the 2-norm of functions on a discrete domain or as the Frobenius norm of two tensors. Finally, the rank of a d-dimensional tensor $T = T(x)$ is defined as in [24]:

Definition 1 *For any tensor $T : \Omega_\xi \longrightarrow \mathbb{R}$ we let*

- *$rank(T) = 1$ if T is the tensor product of d vectors $v^{(1)}, \ldots, v^{(d)}$, i.e. if*

$$T \quad = \quad v^{(1)} \otimes v^{(2)} \otimes \cdots v^{(d)}$$

  for all $x \in \Omega_\xi$, and

- *$rank(T) = r$ if r is the minimal number of rank-1 tensors required to represent T as a linear combination.*

3.2 The Dirac-Frenkel-McLachlan variational principle

The Dirac-Frenkel-McLachlan variational principle is very successfully applied in quantum mechanics to reduce high-dimensional Schrödinger equations to computationally tractable problems; cf. [3, 26]. The underlying idea is to seek an approximation $Y(t) \approx P(t)$ on a manifold $\mathcal{M}$ of suitably chosen basis functions. Propagating the approximation in time requires to assign a unique time-derivative $\dot{Y}$ to any $Y \in \mathcal{M}$. In order to keep $Y(t)$ on the ansatz manifold $\mathcal{M}$ for all times, $\dot{Y}$ has to be an element of the tangent space $T_Y \mathcal{M}$ of $\mathcal{M}$ at Y. This condition, however, excludes the straightforward choice $\partial_t Y = \mathcal{A}Y$ because generally $\mathcal{A}Y \notin T_Y \mathcal{M}$. Therefore, $\partial_t Y$ is defined by minimizing the defect $\|\partial_t Y - \mathcal{A}Y\|_2$, i.e.

$$\partial_t Y \quad = \quad \underset{\delta Y \in T_Y \mathcal{M}}{\arg\min} \|\delta Y - \mathcal{A}Y\|_2. \tag{7}$$

This relation is called the Dirac-Frenkel-McLachlan variational principle. Since the minimum is obtained by an orthogonal projection of $\mathcal{A}Y$ onto the tangent space $T_Y\mathcal{M}$ the condition (7) is equivalent to

$$\left\langle \partial_t Y - \mathcal{A}Y \, , \, \delta Y \right\rangle \;\; = \;\; 0 \qquad \text{for all } \delta Y \in T_Y\mathcal{M}. \tag{8}$$

Note that the solution of (7) usually differs from the best approximation

$$\hat{Y} \;\; = \;\; \operatorname*{argmin}_{Y \in \mathcal{M}} \|Y - P\|_2 \tag{9}$$

of the exact solution P. The advantage of the variational principle is that (7) provides an equation of motion for the approximation Y whereas computing the best approximation $\hat{Y}$ requires explicit knowledge of the exact solution of the CME.

3.3 Low-rank ansatz functions

The quality of the variational approximation Y depends, of course, on the particular choice of the ansatz manifold. In the low-rank approach the ansatz

$$Y(t,x) \;\; = \;\; \sum_{J \leq r} a_J(t) U_J(t,x) \qquad \text{with} \qquad U_J(t,x) = \prod_{i=1}^{d} u_{j_i}^{(i)}(t,x_i) \tag{10}$$

with time-dependent basis functions $u_{j_i}^{(i)} : \mathbb{R} \times \{0, \ldots, \xi_i\} \longrightarrow \mathbb{R}$ and coefficients $a_J(t) \in \mathbb{R}$ are used; cf. [3, 23, 30, 32]. The sum is taken over all multi-indices $J = (j_1, \ldots, j_d)$ such that $1 \leq j_i \leq r_i$ where $r \in \mathbb{N}^d$ is a chosen rank vector. Since the above representation is not unique the orthonormality constraints

$$\left\langle u_j^{(i)} \, , \, u_k^{(i)} \right\rangle \;\; = \;\; \delta(j,k), \qquad \left\langle \partial_t u_j^{(i)} \, , \, u_k^{(i)} \right\rangle \;\; = \;\; 0 \tag{11}$$

are imposed for $i \in \{1, \ldots, d\}$ and $j, k \in \{1, \ldots, r_i\}$. These constraints guarantee uniqueness of the representation of $\partial_t Y$ in the tangent space $T_Y\mathcal{M}$, cf. [26, 23].

The representation (10) can be understood as a Galerkin ansatz with moving basis functions propagated in time. From the perspective of multi-linear algebra U_J is a rank-1 tensor obtained from the tensor product of the vectors $u_{j_1}^{(1)}, \ldots, u_{j_d}^{(d)}$, and since Y is the linear combination of a (small) number of rank-1 tensors the ansatz (10) is called *dynamical low-rank approximation* (DLRA) of P. The idea to choose a basis function of the form (10) and to propagate it via the variational principle was introduced as the *multiconfiguration time-dependent Hartree method* (MCTDH) to approximate Schrödinger equations in many dimensions (see, e.g., [30, 3, 25, 26] and references therein).

3.4 Dynamical low-rank approximation

Let $\mathcal{M}$ be the manifold of all basis functions:

$$\mathcal{M} = \left\{ Y : \Omega_\xi \longrightarrow \mathbb{R} \quad \text{with } Y(x) = \sum_{J \leq r} a_J U_J(x), \right.$$

$$\left. a_J \in \mathbb{R}, \quad U_J(x) = \prod_{i=1}^d u_{j_i}^{(i)}(x_i) \quad \text{and constraints (11)} \right\}. \quad (12)$$

For any $Y \in \mathcal{M}$ the derivative $\partial_t Y$ is defined by (7), but in order to propagate Y, equations of motion for the coefficients a_J and the basis functions $u_{j_i}^{(i)}$ are required. In this section we show how these equations can be obtained from the variational principle.

For any $Y(t, \cdot) \in \mathcal{M}$, $l \in \{1, \ldots, d\}$ and $j \in \{1, \ldots, r_d\}$ let $\bar{Y}_j^{(l)}$ be the function

$$\bar{Y}_j^{(l)}(x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_d) = \left\langle u_j^{(l)}, Y \right\rangle$$

$$= \sum_{x_l} u_j^{(l)}(x_l) Y(x_1, \ldots, x_l, \ldots, x_d).$$

Since $\bar{Y}_j^{(l)}$ depends on all space directions but x_j, these expressions are called *single-hole functions* in the MCTDH approach, cf. [3]. This notation is useful because due to (11) the time-derivative of $Y(t, x)$ can be re-written as

$$\partial_t Y = \sum_{J \leq r} \dot{a}_J U_J + \sum_{i=1}^d \sum_{k=1}^{r_i} \partial_t u_k^{(i)} \cdot \bar{Y}_k^{(i)}. \quad (13)$$

Theorem 1 *Let $M^{(l)}(Y)$ be the symmetric mass matrix with entries*

$$m_{jk}^{(l)}(Y) = \left\langle \bar{Y}_j^{(l)}, \bar{Y}_k^{(l)} \right\rangle, \quad j, k \in \{1, \ldots, r_l\}. \quad (14)$$

Suppose that $M^{(l)}(Y)$ is regular and let $w_{jk}^{(l)}(Y)$ be the entries of the inverse $W^{(l)}(Y) = \left(M^{(l)}(Y) \right)^{-1}$. Then, the equations of motion for the dynamic low rank approximation in terms of its coefficients a_J and basis functions $u_i^{(l)}$ are given by

$$\dot{a}_J = \langle \mathcal{A}Y, U_J \rangle \quad \text{for } J \leq r \quad (15)$$

$$\partial_t u_i^{(l)} = \sum_{k=1}^{r_l} w_{ik}^{(l)}(Y)(I - \mathcal{P}^{(l)}) \left\langle \mathcal{A}Y, \bar{Y}_k^{(l)} \right\rangle \quad \forall l \leq d, \forall i \leq r_l. \quad (16)$$

Here, $\mathcal{P}^{(l)}$ denotes the orthogonal projection onto the space spanned by $\left\{ u_j^{(l)}, j = 1, \ldots, r_l \right\}$:

$$\mathcal{P}^{(l)} v = \sum_{j=0}^{r_l} \left\langle v, u_j^{(l)} \right\rangle u_j^{(l)}.$$

Remark. 1. Since $\bar{Y}_k^{(l)}$ does not depend on x_k, the inner product in Eq. (16) is taken with respect to all x_i except x_l , and since $\mathcal{A}Y$ depends on all variables the result is a function in t and x_l as required.

2. For the implementation, the equations of motion can be formulated in a notationally more complicated, but computationally more efficient way which takes advantage of the tensor structure; see Sec. 3.5.

The theorem is proved by adapting the derivation of the MCTDH equations in [30] to the CME (cf. [23, 32] for the two- and three-dimensional case):

Proof of Theorem 1. For any $Y \in \mathcal{M}$ the elements of the tangent space $T_Y \mathcal{M}$ are of the form

$$\delta Y \;=\; \sum_{J \leq r} \delta a_J U_J + \sum_{i=1}^{d} \sum_{k=1}^{r_i} \delta u_k^{(i)} \bar{Y}_k^{(i)} \tag{17}$$

$$\text{with } \left\langle u_j^{(i)} \,,\, u_k^{(i)} \right\rangle \;=\; \delta(j,k), \quad \left\langle \delta u_j^{(i)} \,,\, u_k^{(i)} \right\rangle = 0.$$

First, choose a multi-index $\tilde{J}$ and consider the variation $\delta Y = \delta a_{\tilde{J}} \cdot U_{\tilde{J}}$, i.e. let $\delta a_K = 0$ for all $K \neq \tilde{J}$ and $\delta u_k^{(i)} = 0$ for all i, k. Inserting this into the variational principle (8) and, for simplicity, omitting the tilde from the multi-index, yields

$$\delta a_J \left\langle \partial_t Y \,,\, U_J \right\rangle \;=\; \delta a_J \left\langle \mathcal{A}Y \,,\, U_J \right\rangle \qquad \text{for all } \delta a_J.$$

Since we have

$$\left\langle \partial_t Y \,,\, U_J \right\rangle \;=\; \sum_{K \leq r} \dot{a}_K \left\langle U_K \,,\, U_J \right\rangle + \sum_{K \leq r} a_K \left\langle \partial_t U_K \,,\, U_J \right\rangle \;=\; \dot{a}_J$$

due to the orthonormality constraints (11), this leads to (15). Next, choose some $l \in \{1, \dots, d\}$ and $\mu \in \{1, \dots, r_l\}$ and let $\delta Y = \delta u_\mu^{(l)} \bar{Y}_\mu^{(l)}$. This is the variation with respect to $\delta u_\mu^{(l)}$ while keeping all other basis functions and all coefficients fixed. For this choice of δY the variational principle (8) reads

$$\left\langle \partial_t Y \,,\, \delta u_\mu^{(l)} \bar{Y}_\mu^{(l)} \right\rangle \;=\; \left\langle \mathcal{A}Y \,,\, \delta u_\mu^{(l)} \bar{Y}_\mu^{(l)} \right\rangle. \tag{18}$$

After substituting (13) and using that the orthonormality constraints imply

$$\left\langle U_J \,,\, \delta u_\mu^{(l)} \right\rangle \;=\; 0 \qquad \text{for all } J$$

$$\left\langle \partial_t u_k^{(i)} \bar{Y}_k^{(i)} \,,\, \delta u_\mu^{(l)} \right\rangle \;=\; 0 \qquad \text{for all } i \neq l$$

we obtain

$$\left\langle \partial_t Y \,,\, \delta u_\mu^{(l)} \bar{Y}_\mu^{(l)} \right\rangle \;=\; \sum_{k=1}^{r_l} \left\langle \partial_t u_k^{(l)} \cdot \bar{Y}_k^{(l)} \,,\, \delta u_\mu^{(l)} \bar{Y}_\mu^{(l)} \right\rangle$$

$$=\; \sum_{k=1}^{r_l} m_{k\mu}^{(l)}(Y) \left\langle \partial_t u_k^{(l)} \,,\, \delta u_\mu^{(l)} \right\rangle.$$

Inserting this into the left-hand side of (18) yields

$$\sum_{k=1}^{r_l} m_{k\mu}^{(l)}(Y) \left\langle \partial_t u_k^{(l)} , \delta u_\mu^{(l)} \right\rangle = \left\langle \mathcal{A}Y , \delta u_\mu^{(l)} \bar{Y}_\mu^{(l)} \right\rangle$$

for all $\delta u_\mu^{(l)}$ with $\left\langle \delta u_\mu^{(l)} , u_j^{(l)} \right\rangle = 0$ and hence

$$\sum_{k=1}^{r_l} m_{k\mu}^{(l)}(Y) \partial_t u_k^{(l)} = (I - \mathcal{P}^{(l)}) \left\langle \mathcal{A}Y , \bar{Y}_\mu^{(l)} \right\rangle.$$

Eq. (16) follows by taking the inverse of the mass matrix $M^{(l)}(Y)$, which was assumed to be regular. ∎

Combining the FSP method with error estimates for the dynamical low-rank approximation we may bound the approximation error of the low-rank approximation $Y(t)$ compared to the exact solution, restricted to the truncated state space Ω_ξ. Let us denote the latter by $P(t)|_\xi$, and furthermore, denote by $\hat{Y}$ the best approximation of $P|_\xi(t)$ for $0 \leq t \leq T$. Recall that $P_\xi(t)$ denotes the exact solution of the truncated CME (see (5)). Then, for $d = 2$ the following theorem holds:

Theorem 2 *Suppose that the finite state projection satisfies $P_\xi(t, \Omega_\xi) \geq 1 - \varepsilon$ and that a continuously differentiable best approximation $\hat{Y} \in \mathcal{M}$ to $P_\xi(t)$ exists for $0 \leq t \leq T$. Let the r-th singular value $\sigma_r(t)$ of $\hat{Y}(t)$ be uniformly bounded away from zero, i.e., $\sigma_r(t) \geq \rho > 0$, and assume that the best approximation error is bounded by $\|P_\xi(t) - \hat{Y}(t)\| \leq \rho/16$ for $0 \leq t \leq T$. Then, there exists a constant $\alpha = \alpha(\varepsilon) > 0$ such that*

$$\max_{x \in \Omega_\xi} |P(t,x) - Y(t,x)| \leq \varepsilon + \rho/16 + \alpha e^{\alpha t} \int_0^T \|P_\xi(s) - \hat{Y}(s)\| ds \quad (19)$$

for $0 \leq t \leq T$, and as long as the integrand is bounded by $\rho/8$.

Proof of Theorem 2. Adding $P_\xi(t) - P_\xi(t) + \hat{Y}(t) - \hat{Y}(t) = 0$, applying the triangular inequality and exploiting the equivalence between norms in finite dimensions ($\| \cdot \|_{\max} \leq \| \cdot \|$) yields:

$$\max_{x \in \Omega_\xi} |P(t,x) - Y(t,x)| \leq \max_{x \in \Omega_\xi} |P(t,x) - P_\xi(t,x)| + $$
$$\|P_\xi(t) - \hat{Y}(t)\| + \|\hat{Y}(t) - Y(t)\|.$$

The first term is bounded by ε based on the FSP method. The second term is uniformly bounded by $\rho/16$ by assumption. Finally, the third term can be bounded by exploiting Theorem 6.1 of [23]: Since the ODE is linear, the bounds (6.8)-(6.10) in [23, Thm. 6.1] hold with $\lambda = L = \|\mathcal{A}\|$, $\alpha = 16\mu/\rho + \|\mathcal{A}\|$ and

$$\mu = \max\{\|\mathcal{A}\hat{Y}(t)\|, \|\mathcal{A}Y(t)\| : 0 \le t \le T\}. \qquad \blacksquare$$

Theorem 2 states that the error between the exact solution of the (full) CME, truncated to the finite state space Ω_ξ, and the DRLA solution of the truncated CME can be made arbitrary small upon increasing the state space and the dimension of the approximation manifold, i.e., the rank of the dynamical low-rank approximation. Hence, Theorem 2 theoretically backs up the approximation properties of the DLRA.

3.5 Efficient implementation

The equations of motion for the low rank approximation (15) and (16) typically contain significantly less degrees of freedom than the CME on the truncated state space. As an example, suppose that the limits ξ_i of the truncated state space Ω_ξ are the same in each direction, i.e. $\xi_1 = \ldots = \xi_d = \eta - 1$ for some $\eta \in \mathbb{N}$. Then, the CME is equivalent to a system of η^d ODEs. In contrast, the DLRA contains only $(r_1 + \ldots + r_d)\eta + (r_1 \cdot \ldots \cdot r_d)$ degrees of freedom because there are $r_1 + \ldots + r_d$ basis functions $u_{j_i}^{(i)}$ each of which is a vector of length η, and because a_J is a tensor with $r_1 \cdot \ldots \cdot r_d$ entries. In cases where $\eta \gg d \ge 2$ and $\eta \gg r_i$ this amounts to a significant reduction of the computational costs. In the example presented in Section 4.2, the DLRA decreases the number of ODEs by more than 99.7 %.

The price to pay for this reduction, however, is the fact that the new equations of motion have a more complicated structure than the CME. First of all, the differential equations (15) and (16) are *nonlinear*. Moreover, (16) contains an inverse mass matrix which becomes ill-conditioned or even singular if the rank of the DLRA is larger than the (essential) rank of the true solution—a problem known as *over-approximation*. It has been investigated in [23], and we will further comment on this in Sec. 5. Finally, the constraints (11) have to be respected during the numerical solution of the differential equations. This can be done, e.g., by applying projection methods (cf. Sec. IV.4 and V.5 in [19]) or using the splitting method from [25, Sec. 5].

The main idea of the DLRA is to propagate only one-dimensional basis functions $u_{j_i}^{(l)}$ and the corresponding coefficients a_J instead of the large, multi-dimensional tensor Y. If Y is explicitly required (e.g. in order to visualise the solution), it can be reconstructed via Eq. (10). However, the efficiency of the DLRA is considerably increased if reconstruction of and computations with the full tensor $Y(t)$ are completely avoided during the time integration of the equations of motion (15) and (16). In this section, we outline how this can be achieved.

The crucial observation is the fact that the propensity functions $\alpha_m(x)$ of the CME (1)

$$\partial_t P(t,x) \quad = \quad \sum_{m=1}^{N} \Big(\alpha_m(x - \nu_m)P(t, x - \nu_m) - \alpha_m(x)P(t,x) \Big)$$

are rank-1 tensors and can therefore be written as

$$\alpha_m(x) \quad = \quad c_m \alpha_m^{(1)}(x_1) \alpha_m^{(2)}(x_2) \cdot \ \ldots \ \cdot \alpha_m^{(d)}(x_d), \tag{20}$$

cf. [8, 12]. For example, in case of the reaction $S_1 + S_2 \xrightarrow{c_m} S_3$ we have $\alpha_m^{(1)}(x_1) = x_1$, $\alpha_m^{(2)}(x_2) = x_2$, and $\alpha_m^{(l)}(x_l) = 1$ for all $l > 2$. Even in more complicated reaction schemes, including inhibition by other species (cf. the numerical examples in Sec. 4), the corresponding propensities may still have a product structure (20).

This allows, for example, to evaluate the inner product $\langle \mathcal{A}Y \, , \, U_J \rangle$, which occurs in Eq. (15), in the following way:

$$
\begin{aligned}
\langle \mathcal{A}Y \, , \, U_J \rangle \quad &= \quad \sum_{m=1}^{N} \sum_{x \in \Omega_\xi} \alpha_m(x - \nu_m) Y(t, x - \nu_m) U_J(t, x) \\
&\quad - \sum_{m=1}^{N} \sum_{x \in \Omega_\xi} \alpha_m(x) Y(t, x) U_J(t, x) \\
&= \quad \sum_{m=1}^{N} c_m \sum_{K \leq r} a_K(t) \left(\prod_{i=1}^{d} \gamma_1(i, m, k_i, j_i) - \prod_{i=1}^{d} \gamma_0(i, m, k_i, j_i) \right)
\end{aligned}
$$

with

$$
\begin{aligned}
\gamma_1(i, m, k_i, j_i) \quad &= \quad \sum_{x_i=0}^{\xi_i} \alpha_m^{(i)}(x_i - \nu_m^{(i)}) u_{k_i}^{(i)}(t, x_i - \nu_m^{(i)}) u_{j_i}^{(i)}(t, x_i) \\
&= \quad \left\langle \alpha_m^{(i)}(\, \cdot \, - \nu_m^{(i)}) u_{k_i}^{(i)}(t, \, \cdot \, - \nu_m^{(i)}) \, , \, u_{j_i}^{(i)}(t, \, \cdot \,) \right\rangle \\
\gamma_0(i, m, k_i, j_i) \quad &= \quad \sum_{x_i=0}^{\xi_i} \alpha_m^{(i)}(x_i) u_{k_i}^{(i)}(t, x_i) u_{j_i}^{(i)}(t, x_i) \\
&= \quad \left\langle \alpha_m^{(i)}(\, \cdot \,) u_{k_i}^{(i)}(t, \, \cdot \,) \, , \, u_{j_i}^{(i)}(t, \, \cdot \,) \right\rangle .
\end{aligned}
$$

Hence, the expensive inner product $\langle \mathcal{A}Y \, , \, U_J \rangle$ with respect to the multi-dimensional variable x can be replaced by inner products γ_1 and γ_0 with respect to single spatial directions x_i, which are much cheaper to compute. In particular, the tensor Y does not have to be reconstructed. Moreover, if the i-th spatial direction is not affected by the m-th reaction, then $\nu_m^{(i)} = 0$, $\alpha_m^{(i)} = 1$ and $\gamma_1(i, m, k_i, j_i) = \gamma_0(i, m, k_i, j_i) = \delta(k_i, j_i)$. This means that in the above expression many terms vanish, which can be exploited to reduce the computational costs. Analogously the product structure (20) allows to compute the terms appearing in the second equation of motion (Eq. (16)) in an efficient way.

4 Numerical examples

4.1 A toggle switch

A genetic toggle consists of a mutually repressing gene pair, where there are two competing proteins, S_1 and S_2, each of which inhibits the transcription of its opponent. Gene toggles are present in different biological systems such as, e.g., bacteriophage-λ. As a model for such a toggle switch the following reaction system with $b = 0.4$ and $c = 0.05$ was considered in [21]:

$$
\begin{array}{|c|ccc|ccc|}
\hline
\text{No.} & \multicolumn{3}{c|}{\text{Reaction}} & \multicolumn{3}{c|}{\text{Propensity function}} \\
\hline
1. & S_1 & \longrightarrow & \star & \alpha_1(x) & = & c \cdot x_1 \\
2. & S_2 & \longrightarrow & \star & \alpha_2(x) & = & c \cdot x_2 \\
3. & \star & \longrightarrow & S_1 & \alpha_3(x) & = & b/(b + x_2) \\
4. & \star & \longrightarrow & S_2 & \alpha_4(x) & = & b/(b + x_1) \\
\hline
\end{array}
\tag{21}
$$

Both proteins S_1 and S_2 can decay via the first two reactions. The transcription of new copies of S_1 is inhibited if the abundance of S_2 is large, because then the propensity of the third reaction almost vanishes, whereas vice versa the production of S_2 by reaction four is inhibited by S_1.

As a first example, the corresponding CME was solved by the DLRA on the time interval [0,500]. As initial distribution, a "discrete Gaussian"

$$
\begin{aligned}
P(0, x) &= \gamma \cdot \exp\left(-0.5(x - \mu)^T C^{-1}(x - \mu)\right) \quad \text{for all } x \in \Omega_\xi, \\
C &= \frac{1}{2}\begin{pmatrix} 75 & -15 \\ -15 & 75 \end{pmatrix}
\end{aligned}
$$

centered at $\mu = (30, 5)$ was considered; the normalization constant γ was determined by the condition $\sum_{x \in \Omega_\xi} P(0, x) = 1$. In this example, the truncated state space $\Omega_{50,50}$ was sufficiently small such that a high-accuracy reference solution of the CME on the truncated state space could be obtained with the MATLAB routine ode45. Fig. 1 shows the DLRA approximation (left column) and the reference solution (right columns) at different times. The DLRA used only five dynamical basis functions in each direction ($r = (5, 5)$). This reduced the number of degrees of freedom from $51^2 = 2601$ (truncated CME) to $2 \cdot 5 \cdot 51 + 5^2 = 535$ (DLRA). The DLRA equations (15) and (16) were integrated with MATLAB's ode45 routine. This method does not exactly respect the orthonormality constraints, but the error was neglible. In spite of the length of the time interval, computing the DLRA solution took only 3 minutes and 2 seconds on a PC with an 3.00 GHz Intel (R) Pentium (R) 4 CPU. Nevertheless, Figures 1 and 2 show that the DLRA results in a very good approximation of the exact solution of the truncated CME. The error of the DLRA solution in the 2-norm is plotted in Fig. 3 along with the error of the best approximation. The best approximation was obtained by computing the singular value decomposition of the reference solution matrix P at each time point and setting all but the first five singular values to zero. Fig. 3 reveals that the DLRA error is only slightly larger

than the error of the best approximation. For two-dimensional problems, this observation is explained by the error estimate given in Theorem 2.

4.2 Lambda phage

In a second and more challenging example, the DLRA was applied to a model for the life cycle of bacteriophage-λ [21]. This model comprises five species $S_1, \ldots, S_5$ and ten reactions:

No.	Reaction	Propensity function
1.	$\star \longrightarrow S_1$	$\alpha_1(x) = a_1 b_1/(b_1 + x_2)$
2.	$\star \longrightarrow S_2$	$\alpha_2(x) = (a_2 + x_5)b_2/(b_2 + x_1)$
3.	$\star \longrightarrow S_3$	$\alpha_3(x) = a_3 b_3 x_2/(b_3 x_2 + 1)$
4.	$\star \longrightarrow S_4$	$\alpha_4(x) = a_4 b_4 x_3/(b_4 x_3 + 1)$
5.	$\star \longrightarrow S_5$	$\alpha_5(x) = a_5 b_5 x_3/(b_5 x_3 + 1)$
6.	$S_1 \longrightarrow \star$	$\alpha_6(x) = c_1 \cdot x_1$
7.	$S_2 \longrightarrow \star$	$\alpha_7(x) = c_2 \cdot x_2$
8.	$S_3 \longrightarrow \star$	$\alpha_8(x) = c_3 \cdot x_3$
9.	$S_4 \longrightarrow \star$	$\alpha_9(x) = c_4 \cdot x_4$
10.	$S_5 \longrightarrow \star$	$\alpha_{10}(x) = c_5 \cdot x_5$

$$(22)$$

The life cycle of bacteriophage-λ consists of two phases: the lysogenic and the lytic phase. The transition from one phase to the other is regulated by a toggle switch of two competing proteins S_1 and S_2 which inhibit each other's transcription, similar to the previous example. The lysogenic phase is characterised by a large abundance of S_2 suppressing the production of S_1 via reaction 1. The generation of new copies of S_2 via reaction 2, however, depends on the amount of S_5. If signs of stress in the host cell are detected, the number of S_5 copies decreases, and the phage enters into the lytic phase: the abundance of S_2 is reduced such that more S_1 copies can be generated via reaction 1 which then inhibit the production of S_2. For further discussion see [21] and references therein.

The solution of the CME corresponding to (22) was approximated by the DLRA on the time interval $[0, 10]$. Parameters were taken from [21]:

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
a_i	0.5	1	0.15	0.3	0.3
b_i	0.12	0.6	1	1	1
c_i	0.0025	0.0007	0.0231	0.01	0.01

The initial distribution was a sharp peak at the origin, namely a multinomial

distribution with parameters $n = 3$ and $p = (0.05, \ldots, 0.05)$:

$$P(0, x) = \begin{cases} \dfrac{3!}{x_1! \; \cdots \; x_5!(3 - |x|)!} 0.05^{|x|}(1 - 5 \cdot 0.05)^{3-|x|} & \text{if } |x| \leq 3 \\ 0 \quad \text{else} \end{cases} \tag{23}$$

where $|x| = x_1 + \ldots + x_5$.

Results were computed on the truncated state space Ω_ξ with truncation index $\xi = (15, 40, 10, 10, 10)$. In each spatial direction, only four basis functions were used (i.e. $r = (4, 4, 4, 4, 4)$) which reduced the number of degrees of freedom from $15 \cdot 40 \cdot 10 \cdot 10 \cdot 10 = 600,000$ (truncated CME) down to $4 \cdot (15 + 40 + 10 + 10 + 10) + 4^5 = 1364 \equiv 0.2273$ % (DLRA). The DLRA equations (15) and (16) were solved with MATLAB's `ode45` routine in 5 minutes and 0 seconds.

The DLRA results were compared with a probability distribution generated by the stochastic simulation algorithm (SSA). $2,500,000$ runs of the SSA were performed, which took 2 hours, 47 minutes, and 40 seconds. The difference between the DLRA and the SSA probability distribution was less than $3.04 \cdot 10^{-3}$. The one-dimensional marginal distributions of the DLRA solution and the SSA solution agree very well as can be seen in Figure 4. The two-dimensional marginal distribution with respect to the S_1-S_2 plane is shown in Figure 5; here, the difference between the DLRA and the SSA approximation is less than $6.84 \cdot 10^{-3}$. It is worth noticing that the SSA approximation is not necessarily more accurate than the DLRA. With 2,500,000 runs of the SSA, the initial distribution (23) was sampled only up to a maximal error of 0.0268774. Hence, the SSA result cannot be considered as an "exact" reference solution. Nevertheless, the comparison of the DLRA with the SSA indicates that the DLRA provides an approximation of comparable accuracy with significantly lower computational cost.

5   Discussion

The previous section demonstrates the capability of the DLRA to turn the CME into a computationally tractable problem by efficiently reducing the number of differential equations. In this section we comment on potential extensions and refinements. Work on these extensions is currently in progress.

Conservation of mass.   It is well known that the full CME is mass-preserving, i.e.,

$$\sum_{x \in \mathbb{N}^d} P(t, x) = \sum_{x \in \mathbb{N}^d} P(0, x) = 1.$$

On the truncated state space Ω_ξ the total mass is still approximately conserved provided that the truncation index is large enough. The variational principle, however, does not inherit this qualitative property of the CME to the DLRA. In some cases this problem can easily be fixed by a renormalisation of $P(t, \cdot)$, but

it would be more advantageous to achieve true mass conservation by inheriting this property to a suitable modified dynamical low rank approximation.

Adaptive rank strategy. In our numerical example the rank vector r was chosen based on the initial distribution and kept constant during the entire computation. Yet the "optimal" rank—being roughly characterized as the minimal rank that ensures a pre-defined tolerance—might change with time. From a numerical point of view, an adaptive rank control is desirable. The choice of r could be based on available information such as the rank of the mass matrices or the defect $\|\partial_t Y - \mathcal{A}Y\|_2$.

Initial data for the DLRA. In many applications, the initial distribution of a CME is assumed to be the discrete delta function $\delta(\cdot, X^0)$ at some initial state X^0. Since the rank of such a delta peak is 1, starting the DLRA with more than one basis function per spatial direction would lead to over-approximation at the very beginning of the time-interval. This problem can be avoided by the following initial step: The CME is solved on a *short* time interval $[0, t_0]$ in a *small* neighborhood of the initial state, say, for all $x \in \Omega_\xi$ with $|x_j - X_j^0| \leq \kappa$ with some small $\kappa \in \mathbb{N}$. The resulting truncated CME contains only $(2\kappa + 1)^d$ ODEs and can therefore be solved with some standard method. The parameter κ should be chosen such that the truncation error is negligible (the FSP provide error estimates), but a rather small value is typically sufficient because the solution remains sharply located for some time. The rank of the solution, however, increases such that initial data for the DLRA can be obtained by truncating the singular value decomposition of $P(t_0, \cdot)$ according to the chosen rank vector r. In a d-dimensional problem with $d > 2$ the multilinear singular value decomposition from [24] can be used. This strategy is generally feasible but computationally rather costly. Eventually, the aim is to solve this problem with the adaptive rank control approach mentioned earlier. An adaptive rank vector r would allow to start the DLRA with a rank-1 configuration and increase the number of basis functions only when necessary.

Summary. In the past years it has become evident that stochastic effects in regulatory networks play an important role, leading to an increasing in stochastic modelling attempts. So far, research has mainly concentrated on *indirectly* approximating the CME by many realizations of the associated Markov process, but we believe that solving the CME *directly* offers not only a feasible alternative, but also a different perspective that might provide new insight into the nature of stochastic reaction kinetics. The results presented in this article indicate the potential of the dynamical low-rank approximation to decrease the complexity of the CME to such an extend that an efficient approximation of the solution up to a acceptable approximation error is possible. Further research in this direction is work in progress.

References

[1] A. Alfonsi, E. Cancès, G. Turinici, B. D. Ventura, and W. Huisinga. Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. *ESAIM Proceeding*, 14:1–13, 2005.

[2] A. P. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics*, 149:1633–1648, 1998.

[3] M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer. The multiconfiguration time-dependent hartree method: A highly efficient algorithm for propagating wavepackets. *Physics Reports*, 324:1–105, 2000.

[4] K. Burrage, M. Hegland, S. MacNamara, and R. B. Sidje. 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 and W.J.Stewart, editors, *Markov Anniversary Meeting: An international conference to celebrate the 150th anniversary of the birth of A.A. Markov*, pages 21 – 38. Boson Books, 2006.

[5] K. Burrage and T. Tian. Poisson Runge-Kutta methods for chemical reaction systems. In Y. L. W. Sun and T. Tang, editors, *Advances in Scientific Computing and Applications*, pages 82–96. Science Press, Beijing/New York, 2004.

[6] K. Burrage, T. Tian, and P. Burrage. A multi-scaled approach for simulating chemical reaction systems. *Progress in Biophysics and Molecular Biology*, 85:217–234, 2004.

[7] Y. Cao, D. Gillespie, and L. Petzold. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.*, 122(1):014116, 2005.

[8] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow. Adaptive diskrete Galerkin methods applied to the chemical master equation. Accepted for publication in *SIAM Journal on Scientific Computing*, 2008.

[9] P. Deuflhard and M. Wulkow. Computational treatment of polyreaction kinetics by orthogonal polynomials of a discrete variable. *IMPACT Comput. Sci. Eng.*, 1(3):269–301, 1989.

[10] W. E, D. Liu and E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *Journal of Chemical Physics*, 123:194107–, 2005.

[11] M. B. Elowitz, E. D. Siggia, P. S. Swain, and A. J. Levine. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[12] S. Engblom. A discrete spectral method for the chemical master equation. Technical Report Technical Report 2006-036, Uppsala University, 2006.

[13] C. W. Gardiner. *Handbook of Stochastic Methods.* Springer, Berlin, 2nd edition edition, 2004.

[14] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.

[15] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[16] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992.

[17] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[18] J. Goutsias. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.*, (122):184102, 2005.

[19] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations.* Number 31 in Springer Series in Computational Mathematics. Springer, 2nd edition edition, 2006.

[20] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics*, 117(15):6959–6969, 2002.

[21] M. Hegland, C. Burden, L. Santoso, S. MacNamara, and H. Booth. A solver for the stochastic master equation applied to gene regulatory networks. *Journal of Computational and Applied Mathematics*, 205:708–724, 2007.

[22] T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54(1):1–26, 2007.

[23] O. Koch and C. Lubich. Dynamical low rank approximation. *SIAM J. Matrix Anal. Appl.*, 29:434–454, 2007.

[24] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278, 2000.

[25] C. Lubich. A variational splitting integrator for quantum molecular dynamics. *Appl. Numer. Math.*, 48(3-4):355–368, 2004.

[26] C. Lubich. On variational approximations in quantum molecular dynamics. *Math. Comput.*, 74(250):765–779, 2005.

[27] S. MacNamara, K. Burrage, and R. B. Sidje. Multiscale modeling of chemical kinetics via the master equation. *SIAM J. Multiscale Modeling and Simulation*, 6(4):1146–1168, 2008.

[28] H. H. McAdams and A. P. Arkin. Stochastic mechanisms in gene expression. *PNAS*, 94:814–819, 1997.

[29] H. H. McAdams and A. P. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet*, 15:65–69, 1999.

[30] H.-D. Meyer, U. Manthe, and L. Cederbaum. The multi-configurational time-dependent hartree approach. *Chem. Phys. Lett.*, 165:73–78, 1990.

[31] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chemical Physics*, 2006.

[32] A. Nonnenmacher and C. Lubich. Dynamical low-rank approximation: applications and numerical experiments. Technical report, University of Tübingen, 2006.

[33] S. Peles, B. Munsky, and M. Khammash. Reduction and solution of the chemical master equation using time-scale separation and finite state projection. *Journal of Chemical Physics*, 125(20):204104, 2006.

[34] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J. Chem. Phys.*, 118(11):4999–5010, 2003.

[35] J. M. Raser and E. K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304:1811–1814, 2004.

[36] M. Rathinam, L. Petzold, Y. Cao, and D. Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *J. Chem. Phys.*, 119:12784–12794, 2003.

[37] M. R. Roussel and R. Zhu. Reducing a chemical master equation by invariant manifold methods. *J. Chem. Phys.*, 121:8716–8730, 2004.

[38] H. Salis and Y. Kaznessis. Accurate hybrid simulation of a system of coupled chemical or biochemical reactions. *J. Chem. Phys.*, 122, 2005.

[39] R. Srivastava, L. You, J. Summers, and J. Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *J. theor. Biol.*, 218:309–321, 2002.

[40] R. Steuer. Effects of stochasticity in models of the cell cycle: from quantized cycle times to noise-induced oscillations. *J Theor Biol*, 228:293–301, 2004.
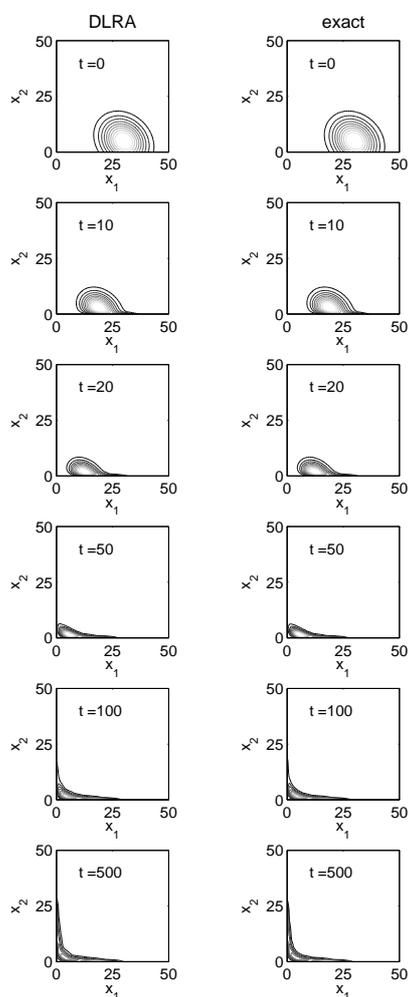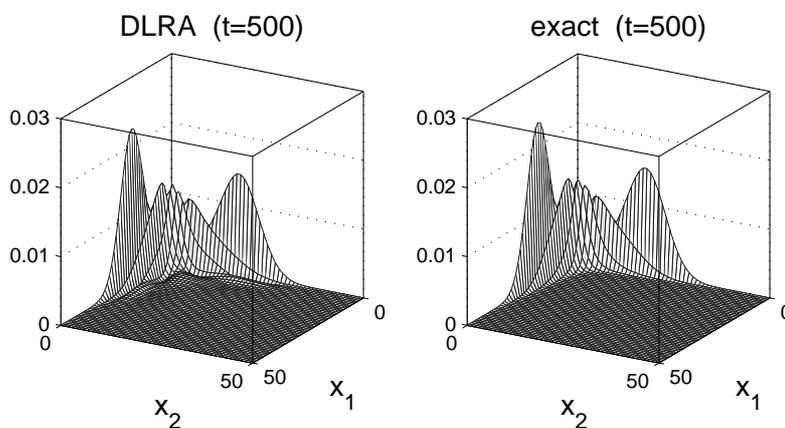
Figure 1: Dynamical low-rank approximation Y (left column) and reference solution P (right column) of the toggle switch at different times. For the sake of clearness, a contour plot based on interpolation is shown though actually both $P(t,x)$ and $Y(t,x)$ are only defined at the discrete grid points $x \in \mathbb{N}^2$. The reference solution was obtained by solving the CME directly with MATLAB's `ode45` and an error tolerance of 1e-10. For the DLRA five ansatz function in each direction were used ($r = (5, 5)$).

Figure 2: Dynamical low-rank approximation Y (left column) and reference solution P (right column) of the toggle switch as mesh plots ($t = 500$). For better visualisation, the plots have been rotated.
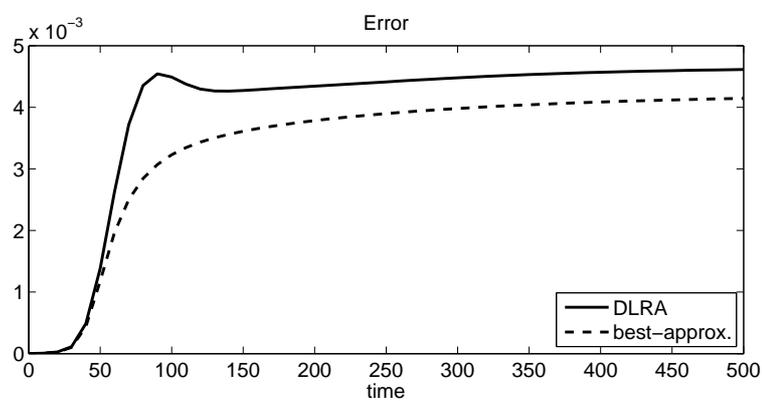

Figure 3: 2-norm error of the DLRA (solid) and of the the best approximation (dashed) of the toggle switch problem as functions in time ($r = (5,5)$). The best approximation is provided by truncating all but the first five singular values of the reference solution matrix.
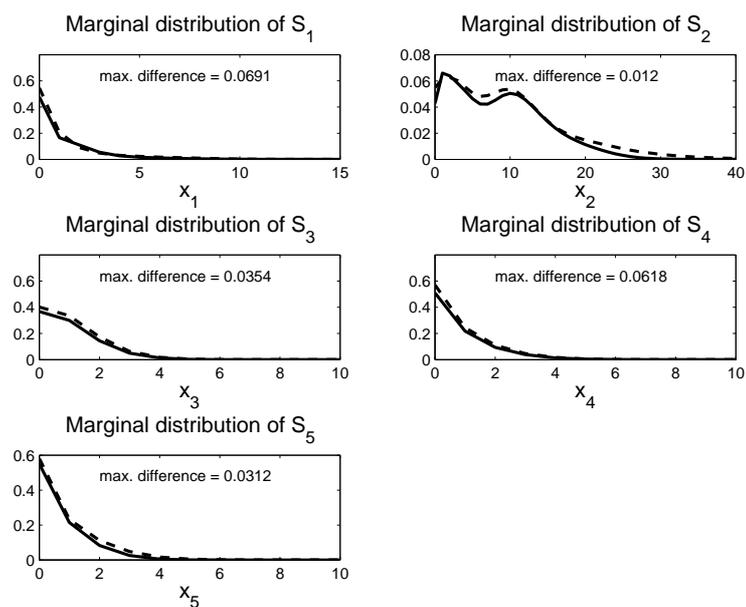
Figure 4: One-dimensional marginal distributions of the solution of the lambda phage model at time $t = 10$. The solid line was computed with the DLRA whereas the dashed line was obtained by 2,500,000 realisations of the SSA. For the DLRA only four ansatz function in each direction were used ($r = (4, 4, 4, 4, 4)$). This reduces the complexity of the problem from $600,000$ down to 1364 ODEs (0.2273 %).
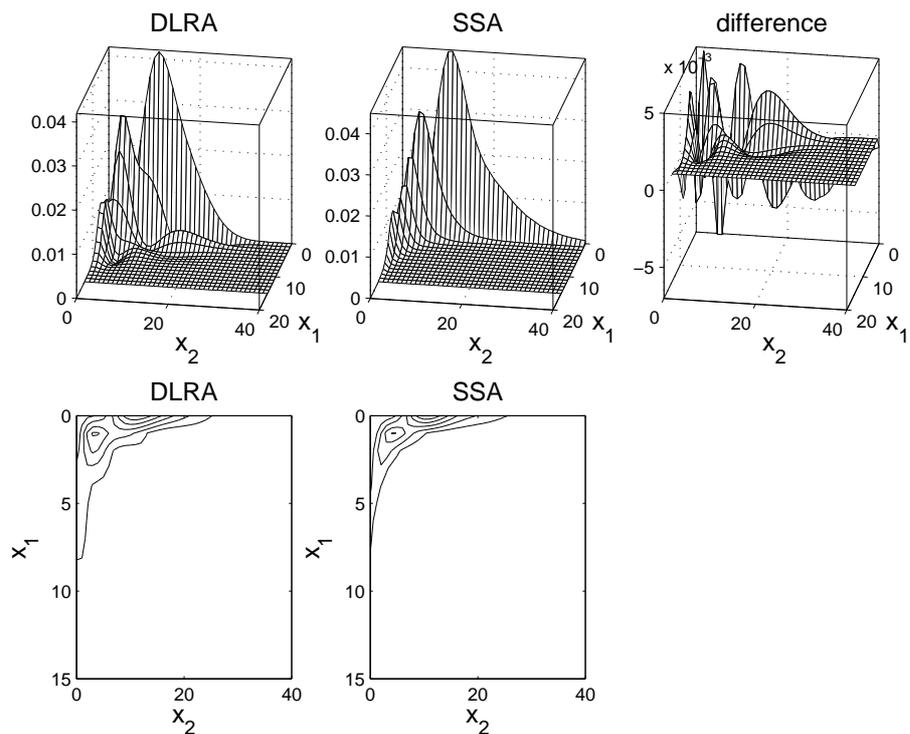
Figure 5: Upper panels: Two dimensional marginal distribution of the S_1-S_2 plane of the lambda phage model. From left to right: DLRA approximation, SSA approximation, difference. Lower panels: contour plots of the DLRA approximation (left) and the SSA approximation (right). For better visualisation the plots have been rotated.