ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Direct Solution of the Chemical Master Equation using Quantized Tensor Trains

V. Kazeev and M. Khammash and M. Nip and C. Schwab

Research Report No. 2013-04
February 2013

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

- Funding SNF: ECCS-0835847
- Funding ERC: AdG247277
- Funding: Human Frontier Science Program Grant RGP0061/2011

Direct Solution of the Chemical Master Equation using Quantized Tensor Trains*

Vladimir Kazeev† Mustafa Khammash‡ Michael Nip§ Christoph Schwab†

February 4, 2013

Abstract

The Chemical Master Equation (CME) is a cornerstone of stochastic analysis and simulation of models of biochemical reaction networks. Yet direct solutions of the CME have remained elusive. Although several approaches overcome the infinite dimensional nature of the CME through projections or other means, a common feature of proposed approaches is their susceptibility to the curse of dimensionality, i.e. the exponential growth in memory and computational requirements in the number of problem dimensions. We present a novel approach that has the potential to "lift" this curse of dimensionality. The approach is based on the use of the recently proposed Quantized Tensor Train (QTT) formatted numerical linear algebra for the low parametric, numerical representation of tensors. The QTT decomposition admits both, algorithms for basic tensor arithmetics with complexity scaling linearly in the dimension (number of species) and sub-linearly in the mode size (maximum copy number), and a numerical tensor rounding procedure which is stable and quasi-optimal. We show how the CME can be represented in QTT format, then use the exponentially-converging *hp*-discontinuous Galerkin discretization in time to reduce the CME evolution problem to a set of QTT-structured linear equations to be solved at each time step using an algorithm based on Density Matrix Renormalization Group (DMRG) methods from quantum chemistry. Our method automatically adapts the "basis" of the solution at every time step guaranteeing that it is large enough to capture the dynamics of interest but no larger than necessary, as this would increase the computational complexity. Our approach is demonstrated by applying it to three different examples from systems biology: independent birth-death process, an example of enzymatic futile cycle, and a stochastic switch model. The numerical results on these examples demonstrate that the proposed QTT method achieves dramatic speedups and 10 to 30 orders of magnitude storage savings over direct approaches.

Keywords: Chemical Master Equation, stochastic models, low rank, tensor approximation, Tensor Train, Quantized Tensor Train, multilinear algebra, mass-action kinetics, stationary distribution.

1 Introduction

In spite of the success of continuous-variable deterministic models in describing many biological phenomena, discrete stochastic models are often necessary to describe biological phenomena inside living cells where random motion of reacting species introduces randomness in both the order and timing of biochemical reactions. Such random effects become more pronounced when one factors in the discrete nature of reactants and the fact that they are often found in low

*Partially supported by the European Research Council (ERC) FP7 programme project AdG247277, NSF Grant ECCS-0835847, and the Human Frontier Science Program Grant RGP0061/2011.

†Seminar für Angewandte Mathematik, ETH Zürich. Rämistrasse 101, 8092 Zürich, Switzerland. {vladimir.kazeev,christoph.schwab}@sam.math.ethz.ch.

‡Department of Biosystems Science and Engineering, ETH Zürich. Mattenstrasse 26, 4058 Basel, Switzerland. mustafa.khammash@bsse.ethz.ch.

§Department of Mechanical Engineering, UC Santa Barbara. Engineering Building II, Santa Barbara, CA 93106-5070. mdnip@engineering.ucsb.edu.

copy numbers inside the cell. Manifestations of randomness vary from copy-number fluctuations among genetically identical cells [1] to dramatically different cell fate decisions [2] leading to phenotypic differentiation within a clonal population. Characterizing and quantifying the effect of stochasticity and its role in the function of cells is a central problem in molecular systems biology.

To account for the random nature of chemical reactions, the evolution of reacting species within living cells is often modeled as a stochastic process. These mathematical models are specified by jump Markov processes where each state represents the population count of each of the constituent species [3]. In this framework, the evolution of the probability density of the system's chemical populations is governed by the Forward Kolmogorov Equation, commonly referred to as the Chemical Master Equation (CME) in the chemical literature. In most cases the CME cannot be solved explicitly and various Monte Carlo simulation techniques have been used to find approximations of the probability densities by producing either detailed or approximate realizations of each process [3, 4, 5]. However, for many systems, biologically important events may occur rarely, necessitating the generation of a prohibitively large set of realizations to obtain sufficiently precise statistics.

At the same time, various approximation methods have been developed that trade accurate density information for computational tractability, often replacing the discrete state-space description with a continuous one. These include Van Kampen's Linear Noise Approximation (LNA) [6], Moment Closure methods [7, 8], and Chemical Langevin Equation (CLE) treatments [9, 10]. These methods tend to give an accurate description of the dynamics when the population counts of all species remain large, but can perform poorly even when a single species exhibits low molecular counts. This is a significant limitation when one needs to model the (boolean) activation state of genes that necessarily have low molecular counts.

The classes of methods described above are complementary and recently there has been much effort attempting to combine the best features of these, leading to the so called hybrid approaches. Many are based on exploiting a time-scale separation to partition the system into subsets of fast and slow reactions and then impose a quasi-stationary assumption to reduce the number of degrees of freedom. These methods are then based on coupling an approximate method such as τ-leaping [11] or the Chemical Langevin Equation [12, 13] for the fast species with an efficient variant of the Gillespie algorithm for the slow species to produce a new Monte Carlo algorithm. Other methods are based on partitioning the chemical species into a subset with large average molecule count and a subset with low molecule count and making an ODE approximation for the dynamics of species with large copy numbers [14, 15]. While these methods result in faster simulations, such speedups come at the cost of accuracy, as modeling errors are introduced by the partial replacement of the CME with cruder descriptions.

Other approaches have attempted to solve the CME directly to obtain the evolution of the probability densities [16], though these analytical solutions apply only to special structures. Alternatively, methods like the finite state projection [17] and the sliding window abstraction [18] are based on truncating the state space to a finite subset containing the majority of the probability mass. These methods have the advantage of providing explicit error bounds on their approximations of the densities. Unfortunately, to guarantee that such an approximation has a low error, it is often necessary to include a large number of states in the truncation, rendering many systems computationally intractable as both storage requirements and computation time become prohibitive.

In order to address this issue, several numerical techniques for compressing the dynamics and the solution have been explored in the recent literature. Attempts were made to expand the probability distribution as a linear combination of a small set of so-called "principal", orthogonal basis functions [19, 20, 21, 22]. Then, either a Galerkin projection was used to map the dynamics onto the lower dimensional subspace spanned by the basis functions (Method of Lines) or first a time discretization was used and then the basis at each time step was adapted by either adding or subtracting basis elements (Rothe's Method). These methods differ primarily in their choice of orthogonal basis. A common feature of these approaches is that they begin with a basis for

probability distributions of a single variable and then use the corresponding tensor product basis for multivariate distributions. This means that they are susceptible to the so-called *curse of dimensionality* [23], that is, the memory requirements and computational complexity of basic arithmetics grow exponentially in the number of dimensions. In the context of the CME, this means that all of these approaches can exhibit an exponential scaling of the complexity with the number of chemical species in the model.

Recent papers have attempted to address the curse of dimensionality by using a low-parametric representation of tensors known as *canonical polyadic* decomposition or *CANDECOMP/PARAFAC*, both notions being subsumed under the acronym *CP* [24, 25]. CP is a methodology for generalizing the singular value decomposition (SVD) for matrices to tensors of dimension greater than two by representing the solution as sums of rank-one tensors (equivalently, linear combinations of distributions in which species counts are independent at each fixed time point). As long as the tensor rank of the solution to be approximated remains low, these approaches can be very computationally efficient as basic arithmetics for tensors in the CP format scales linearly in the number of tensor dimensions.

A key challenge in applying the CP decomposition to construct approximate CME solvers is to control the tensor rank of the computed solution. Basic algebraic tensor operations such as addition and matrix-vector multiplication generally increase rank and hence computational cost. In [26] it is suggested to recompute a lower rank CP decomposition after *every* arithmetic operation. This approach turned out to be problematic in practice. One reason is that the problem of tensor approximation (in the Frobenius norm) with a tensor of fixed rank is, in general, ill-posed [27]. Thus, the numerical algorithms for computing an approximate representation may easily fail. Another reason is that the problem is NP-hard [28, 29] and there is no robust algorithm having any affordable complexity.

Another approach [30], related to the present work, attempts to avoid the problem of approximation in the CP format entirely by projecting the dynamics onto a manifold composed of all tensors with a CP decomposition of some predetermined maximal tensor rank. This procedure results in a set of coupled nonlinear differential equations which are then solved using available ODE solvers. While this effectively controls the tensor rank of the approximate solution, still, to the authors' knowledge, there is no way to estimate either theoretically (*a priori*) or numerically (*a posteriori*) the CP rank of the full CME solution as a function of given data.

In this paper we propose a new, deterministic computational methodology for the direct numerical solution of the CME, without modelling or asymptotic simplifications. The approach has complexity that scales favorably in terms of the number of different species considered and the maximum allowable copy number of each of these species. It is based on the recently proposed *Quantized Tensor Train* (QTT) formatted, numerical tensor algebra [31, 32, 33, 34] which operates on low-parametric, numerical representation of tensors, rather than on their CP representations. This decomposition admits both algorithms for basic tensor arithmetics that scale linearly in the dimension (the species number) and a robust adaptive numerical procedure for the tensor truncation, which is quasi-optimal in the Frobenius norm.

We show in the present paper how the CME can be represented in QTT format, then use *hp*-discontinuous Galerkin discretization in time to exploit the time-analyticity of the CME evolution and to reduce the CME evolution problem to a set of QTT structured linear equations that are solved at each time step [35]. We then exploit an algorithm available for solving linear systems in this format that is based on Density Matrix Renormalization Group (DMRG) methods from quantum chemistry.

The numerical experiments reported below (see, in particular, Table 1) show a 10 to 30 order of magnitude memory savings, which is typically afforded by the new approach presented here.

2 Results and Discussion

A "well-stirred" solution of d chemically reacting molecules in thermal equilibrium can be described by a jump Markov process, where for each fixed time $t \geq 0$, $X(t) \in \mathbb{Z}^d_{\geq 0}$ is a random vector of nonnegative integers with each component representing the number of molecules of one chemical species present in the system. In [6] and the references therein, it is shown that, given an initial condition $X(0) \in \mathbb{Z}^d_{\geq 0}$, the corresponding probability density function (PDF) $\mathbb{Z}^d_{\geq 0} \times [0, \infty) \ni (\underline{x}, t) \mapsto \boldsymbol{p}_{\underline{x}}(t)$ of the process solves the Chemical Master Equation (CME):

$$\frac{\mathrm{d}}{\mathrm{d}t}\, \boldsymbol{p}_{\underline{x}}(t) = -\boldsymbol{p}_{\underline{x}}(t) \sum_{s=1}^{R} \omega^s(\underline{x}) + \sum_{s=1}^{R} \boldsymbol{p}_{\underline{x}-\underline{\eta}^s}(t)\, \omega^s\big(\underline{x} - \underline{\eta}^s\big) \tag{1}$$

where R is the number of reactions in the system, $\underline{\eta}^s \in \mathbb{Z}^d$ and ω^s are the stoichiometric vector and propensity function of the sth reaction, respectively. The CME is a system of coupled linear ordinary differential equations with one equation for each state $X(t) = \underline{x} \in \mathbb{Z}^d_{\geq 0}$.

2.1 Separability and Finite State Projection of the CME operator

Munsky and Khammash [17] rewrote the right-hand side of the CME (1) as the action of a linear operator $\boldsymbol{A}$ on the probability density at the current time:

$$\frac{\mathrm{d}}{\mathrm{d}t}\, \boldsymbol{p}(t) = \boldsymbol{A}\boldsymbol{p}(t) \tag{2}$$

Throughout this paper, we refer to $\boldsymbol{A}$ as the *CME operator*.

Hegland and Garcke introduced an explicit representation of the CME operator as sums and compositions of a few elementary linear operators [26]: let $\boldsymbol{S}_{\underline{\eta}}$ be the spatial shift of a probability density by a vector $\underline{\eta} \in \mathbb{Z}^d$:

$$\Big(\boldsymbol{S}_{\underline{\eta}}\, \boldsymbol{p}\Big)_{\underline{x}} = \boldsymbol{p}_{\underline{x}-\underline{\eta}};$$

and let $\boldsymbol{M}_{\omega}$ be multiplication by a real-valued function ω:

$$\big(\boldsymbol{M}_{\omega}\, \boldsymbol{p}\big)_{\underline{x}} = \omega(\underline{x}) \cdot \boldsymbol{p}_{\underline{x}}\ .$$

Then the CME operator can be written as follows, with $\mathbb{I}$ denoting the identity operator:

$$\boldsymbol{A} = \sum_{s=1}^{R} \Big(\boldsymbol{S}_{\underline{\eta}^s} - \mathbb{I}\Big) \circ \boldsymbol{M}_{\omega^s}\ . \tag{3}$$

To simplify the exposition, we assume that all propensity functions are *rank-one separable*, i.e. they are of the form

$$\omega^s(\underline{x}) = \prod_{k=1}^{d} \omega_k^s(x_k)\,, \quad \underline{x} \in \mathbb{Z}^d_{\geq 0}, \tag{4}$$

for $1 \leq s \leq R$, where each $\omega_k^s(x_k)$ is a nonnegative function in the single variable x_k. Considering rank-one separable propensity functions is sufficient for all elementary reactions which occur as building blocks in more complicated reaction kinetics. We hasten to add, however, that the methods developed herein apply also to models with nonseparable propensities $\omega^s(\underline{x})$.

The CME (2) is posed on the (countably) infinite dimensional space $\mathbb{Z}^d_{\geq 0}$ of states. In this form, the CME (1) is an infinite-dimensional coupled evolution problem which necessitates truncation prior to numerical discretization. In the case of a particular class of monomolecular reactions, Jahnke and Huisinga were able to construct an explicit solution in terms of convolutions of products of Poisson and multinomial distributions [16]. In order to be able to address more complex systems computationally, Munsky and Khammash proposed the Finite State Projection Algorithm (FSP) [17] which seeks to truncate the countably infinite dimensional space $\mathbb{Z}^d_{\geq 0}$ of states of the process to a finite subset over which the dynamics are close to those of the original system.

Theorem 2.1 (Finite State Projection, Theorem 2.2 in [17]). *Consider a Markov process with state space $\mathbb{Z}^d_{\geq 0}$ whose probability density evolves according to the initial value ODE: given an initial state $\boldsymbol{p}_0 \in [0,1]^{\mathbb{Z}^d_{\geq 0}}$, find $\boldsymbol{p}(t) \in [0,1]^{\mathbb{Z}^d_{\geq 0}}$ such that*

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p}(t) = \boldsymbol{A}\boldsymbol{p}(t) \quad 0 \leq t < \infty\,, \qquad \boldsymbol{p}(0) = \boldsymbol{p}_0$$

where the CME operator $\boldsymbol{A} : [0,1]^{\mathbb{Z}^d_{\geq 0}} \mapsto [0,1]^{\mathbb{Z}^d_{\geq 0}}$ can be represented as bi-infinite matrix with nonnegative off-diagonal entries $A_{\underline{x}\underline{x}'}$ indexed by pairs of states $\underline{x}, \underline{x}' \in \mathbb{Z}^d_{\geq 0}$.

With a multi-index $\underline{n} = (n_1, n_2, ..., n_d) \in \mathbb{N}^d$ associate the finite set $\Omega^{\underline{n}}$ of states

$$\Omega^{\underline{n}} = \left\{\underline{x} \in \mathbb{Z}^d_{\geq 0} : 0 \leq x_k \leq n_k - 1 \quad \textit{for} \quad 1 \leq k \leq d\right\} \subset \mathbb{Z}^d_{\geq 0}\,.$$

Let $\boldsymbol{A}^{\underline{n}}$ denote the restriction of $\boldsymbol{A}$ to $\Omega^{\underline{n}}$ and assume that $\boldsymbol{p}_0$ is supported in $\Omega^{\underline{n}}$, i.e. that $\boldsymbol{p}_0 = 0$ in $\mathbb{Z}^d_{\geq 0}\backslash\Omega^{\underline{n}}$. Denote by $\hat{\boldsymbol{p}}_\cdot(\cdot) \in [0,1]^{\Omega^{\underline{n}}}$ the solution of the truncated system with dynamics given by the linear ODE:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\boldsymbol{p}}(t) = \boldsymbol{A}^{\underline{n}}\hat{\boldsymbol{p}}(t)\,, \qquad 0 \leq t < \infty \tag{5}$$

with initial condition $\hat{\boldsymbol{p}}_{\underline{x}}(0) = \boldsymbol{p}_{\underline{x}}(0) = \boldsymbol{p}_0(\underline{x})$. If for some $\epsilon > 0$ and $\tau \geq 0$

$$\sum_{\underline{x} \in \Omega^{\underline{n}}} \hat{\boldsymbol{p}}_{\underline{x}}(\tau) \geq 1 - \epsilon \tag{6}$$

then

$$\hat{\boldsymbol{p}}_{\underline{x}}(\tau) \leq \boldsymbol{p}_{\underline{x}}(\tau) \leq \hat{\boldsymbol{p}}_{\underline{x}}(\tau) + \epsilon \tag{7}$$

for every $\underline{x} \in \Omega^{\underline{n}}$.

Assume that a truncation satisfying (6) can be found, then (7) gives an explicit certificate of the accuracy of the approximate solution. In practice, the truncation required to satisfy a given error tolerance may still require a very large number of states: the dimension of the FSP vector $\hat{\boldsymbol{p}}$ equals $\mathrm{card}(\Omega^{\underline{n}}) = \prod_{k=1}^d n_k$ rendering a direct numerical solution of even the projected equation (5) infeasible in many cases. The remainder of the paper presents a novel approach for the numerical solution of such FSP truncated systems that retain large numbers of states. For notational convenience, we drop the superscripts $\underline{n}$ and the hat from $\hat{\boldsymbol{p}}$ indicating the FSP since we will only consider systems which have already been truncated. Similarly, we now use the shift and multiplication operators in (3) restricted to the truncated state space without change of notation.

Assuming that a FSP has been performed, we henceforth treat $\boldsymbol{p}_{\underline{x}}(t)$ as a d-dimensional $n_1 \times \ldots \times n_d$-vector, i.e. as an array indexed by $\Omega^{\underline{n}}$ which we identify with *ordered* d-tuples of indices $i_k \in \{0, 1, 2, \ldots, n_k - 1\}$, where k ranges from 1 to d. Each dimension k (alternatively referred to as a *mode* or *level*) has a corresponding *mode size* n_k, that is, the number of values which the index for that dimension can take. For our chemically reacting system, $n_k - 1$ corresponds to the maximum number of copies of the kth species that is considered. For a more detailed introduction to basic tensor operations and terminology see, for example, [36, 37].

For the same ordering of $\underline{x}$, consider the corresponding d-dimensional $n_1 \times \ldots \times n_d$-vectors $\boldsymbol{\omega}^s$, $1 \leq s \leq R$, containing the values of the propensities on $\Omega^{\underline{n}}$ to which we shall refer as *propensity vectors*:

$$\boldsymbol{\omega}^s{}_{\underline{x}} = \omega^s(\underline{x}) \quad \text{for all} \quad \underline{x} \in \Omega^{\underline{n}}\,. \tag{8}$$

Within the projected CME (5), the operators corresponding to weighting by the propensity functions, involved in (3), are finite matrices: $\boldsymbol{M}_{\omega^s} = \mathrm{diag}\,\boldsymbol{\omega}^s$. Then, under the rank one separability assumption (4), with $(\boldsymbol{\omega}^s_k)_{x_k} = \omega^s_k(x_k)$ for $0 \leq x_k \leq n_k$, $1 \leq k \leq d$ there holds

$$\boldsymbol{\omega}^s = \boldsymbol{\omega}^s_1 \otimes \ldots \otimes \boldsymbol{\omega}^s_d\,, \qquad 1 \leq s \leq R\,. \tag{9}$$

2.2 The CME in the TT and QTT formats

2.2.1 Tensor Train representation of vectors and matrices

Our approach to the direct numerical solution of the CME (2) is based on the structured, low-parametric representation of all vectors and matrices involved in the solution in the *Tensor Train* (TT) format [31, 38] developed by Oseledets and Tyrtyshnikov. To present it, let us consider a d-dimensional $n_1 \times \ldots \times n_d$-vector $\boldsymbol{p}$ and assume that two- and three-dimensional arrays $U_1, U_2, \ldots, U_d$ satisfy

$$\boldsymbol{p}_{j_1,\ldots,j_d} = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{d-1}=1}^{r_{d-1}} U_1(j_1,\alpha_1) \cdot U_2(\alpha_1, j_2, \alpha_2) \cdot \ldots \cdot U_{d-1}(\alpha_{d-2}, j_{d-1}, \alpha_{d-1}) \cdot U_d(\alpha_{d-1}, j_d) \tag{10}$$

for $0 \le j_k \le n_k - 1$, where $1 \le k \le d$. Then $\boldsymbol{p}$ is said to be represented in the TT decomposition in terms of the *core tensors* $U_1, U_2, \ldots, U_d$. The summation indices $\alpha_1, \ldots, \alpha_{d-1}$ and limits $r_1, \ldots, r_{d-1}$ on the right-hand side of (10) are called, respectively, *rank indices* and *ranks* of the decomposition. Unlike CP, the TT format allows the construction of a decomposition, exact or *approximate*, through the low-rank representation of a sequence of single matrices; for example, by SVD. In particular, note that for every $k = 1, \ldots, d-1$ the decomposition (10) implies a rank-r_k representation of an *unfolding matrix* $\boldsymbol{U}^{(k)}$ which consists of the entries

$$\boldsymbol{U}^{(k)}{}_{\overline{j_1,\ldots,j_k};\,\overline{j_{k+1},\ldots,j_d}} = \boldsymbol{p}_{j_1,\ldots,j_k,j_{k+1},\ldots,j_d}.$$

Conversely, if the vector $\boldsymbol{p}$ is such that the unfolding matrices $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(d-1)}$ are of ranks $r_1, \ldots, r_{d-1}$ respectively, then the cores $U_1, U_2, \ldots, U_d$, such that (10) holds, do exist; see Theorem 2.1 in [38]. The ranks of the unfolding matrices are the lowest possible ranks of a TT decomposition of the vector and, therefore, are called *TT ranks of the vector*.

Example 2.2 (Unfolding of a tensor)**.** *Consider a tensor $\boldsymbol{p}$ of size $3 \times 2 \times 2$. It has two unfolding matrices $\boldsymbol{U}^{(1)}$ and $\boldsymbol{U}^{(2)}$ given by*

$$\boldsymbol{U}^{(1)} = \begin{pmatrix} \boldsymbol{p}_{111} & \boldsymbol{p}_{112} \\ \boldsymbol{p}_{211} & \boldsymbol{p}_{212} \\ \boldsymbol{p}_{311} & \boldsymbol{p}_{312} \\ \boldsymbol{p}_{121} & \boldsymbol{p}_{122} \\ \boldsymbol{p}_{221} & \boldsymbol{p}_{222} \\ \boldsymbol{p}_{321} & \boldsymbol{p}_{322} \end{pmatrix} \quad \textit{and} \quad \boldsymbol{U}^{(2)} = \begin{pmatrix} \boldsymbol{p}_{111} & \boldsymbol{p}_{121} & \boldsymbol{p}_{112} & \boldsymbol{p}_{122} \\ \boldsymbol{p}_{211} & \boldsymbol{p}_{221} & \boldsymbol{p}_{212} & \boldsymbol{p}_{222} \\ \boldsymbol{p}_{311} & \boldsymbol{p}_{321} & \boldsymbol{p}_{312} & \boldsymbol{p}_{322} \end{pmatrix}.$$

While $\boldsymbol{p}$, $\boldsymbol{U}^{(1)}$, and $\boldsymbol{U}^{(2)}$ are structured differently, all have the same entries and represent the same data. The two TT ranks of $\boldsymbol{p}$ are exactly the (matrix) ranks of $\boldsymbol{U}^{(1)}$ and $\boldsymbol{U}^{(2)}$.

Another, fundamental, property of the TT representation is that if the unfolding matrices can be approximated with ranks $r_1, \ldots, r_{d-1}$ and accuracies $\varepsilon_1, \ldots, \varepsilon_{d-1}$ in the Frobenius norm, then the vector itself can be approximated in the TT format with ranks $r_1, \ldots, r_{d-1}$ and accuracy $\sqrt{\sum_{k=1}^{d-1} \varepsilon_k^2}$ in the same norm, which yields a robust and efficient algorithm for the *numerical* low-rank TT approximation of vectors given in full format or in the TT format with higher ranks. For details refer to Theorem 2.2, corollaries and to Algorithms 1 and 2 in [38]. Note also that, unlike CP, the TT representation relies on a certain ordering of the dimensions so that *reordering dimensions may affect the numerical values of TT ranks significantly.* We discuss this issue in Section 2.2.4. Here we note only that the CP decomposition can be considered as a particular case of TT. As for the general TT decomposition (with non-diagonal cores), reordering dimensions may affect TT ranks significantly.

The TT representation may be applied to multidimensional matrices in a similar way as to vectors. Consider a d-dimensional $(m_1 \times \ldots \times m_d) \times (n_1 \times \ldots \times n_d)$-matrix $\boldsymbol{A}$. Let us

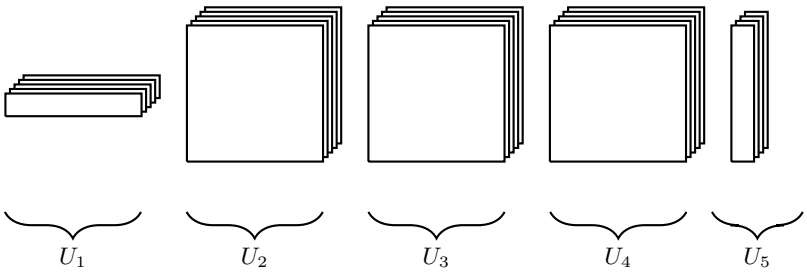


Figure 1: Schematic drawing of a TT decomposition of a five-dimensional array. Each TT core can be visualized as a stack of matrices with the size of the stack equal to the corresponding mode size. The number of TT cores is equal to the number of dimensions of the array. Element $\boldsymbol{u}(j_1, \ldots, j_5)$ of the full array is given by the (matrix) product of matrix j_1 selected from core U_1, matrix j_2 from core U_2, etc. Note that the size of each matrix within a core must be the same, but may differ between distinct cores. Note also that the number of matrices in each core depends on the corresponding mode size of the full tensor and generally differs between cores. Such an interpretation in the sense of a product of parametric matrices is widely used for the *Matrix Product States*, see [39, 40, 41]

vectorize it and merge the corresponding row and column indices to obtain a d-dimensional $m_1 \cdot n_1 \times \ldots \times m_d \cdot n_d$-vector $\boldsymbol{a}$. Then the TT representation of the vector $\boldsymbol{a}$, given by the elementwise equality

$$\boldsymbol{A}_{\substack{i_1,\ldots,i_d \\ j_1,\ldots,j_d}} = \boldsymbol{a}_{\overline{i_1,j_1},\ldots,\overline{i_d,j_d}} = \sum_{\alpha_1=1}^{r_1} \ldots \sum_{\alpha_{d-1}=1}^{r_{d-1}} V_1(i_1, j_1, \alpha_1)$$
$$\cdot\, V_2(\alpha_1, i_2, j_2, \alpha_2) \cdot \ldots \cdot V_{d-1}(\alpha_{d-2}, i_{d-1}, j_{d-1}, \alpha_{d-1}) \cdot V_d(\alpha_{d-1}, i_d, j_d)\,, \tag{11}$$

is called a TT representation of the matrix $\boldsymbol{A}$, the cores $V_1, \ldots, V_d$ are now three- and four-dimensional arrays. Our discussion of the efficiency and robustness of the TT decomposition of vectors also applies to the matrix case.

Note that the *Hierarchical Tensor Representation* [42, 43] itself and coupled with the *tensorization* [44], an extensive overview of which is available in [37], are closely related counterparts of the TT and QTT formats respectively. Also, the structure called now TT decomposition has been known in theoretical chemistry as *Matrix Product States (MPS)*. It has been exploited by physicists to describe quantum spin systems theoretically and numerically for at least two decades now, see [39, 40], cf. [41].

Basic operations of the numerical calculus with vectors and matrices in the TT format, such as addition, Hadamard and dot products, multi-dimensional contraction, matrix-vector multiplication, etc. are considered in detail in [38]. Since the main aim of using tensor-structured approximations is to reduce the complexity of computations and avoid the curse of dimensionality, we emphasize that the storage cost and complexity of basic operations of the TT arithmetics, applied to the representation (10), can be bounded by dnr^α with $\alpha \in \{2, 3\}$, where $n \geq n_1, \ldots, n_d$ and $r \geq r_1, \ldots, r_{d-1}$. This estimate is formally linear in d; however, the TT ranks $r_1, \ldots, r_{d-1}$ in (10) may depend on d and n. Showing that the TT ranks are moderate, e. g. constant or growing linearly with respect to d and constant or growing logarithmically with respect to n, is a crucial issue in the context of TT-structured methods and has been addressed so far mostly experimentally, see, e. g. [45, 46, 47, 48, 49].

Since a TT decomposition of a d-dimensional tensor has $d-1$ ranks that may take different values, it is convenient to introduce an aggregate characteristic such as the *effective rank* of the TT decomposition. For an $n_1 \times \ldots \times n_d$-tensor given in a TT decomposition of ranks $r_1, \ldots, r_{d-1}$, we definite it as the positive root $r_{\text{eff}} = r$ of the quadratic equation

$$n_1 r_1 + \sum_{k=2}^{d-1} r_{k-1} n_k r_k + r_{d-1} n_d = n_1 r + \sum_{k=2}^{d-1} r\, n_k r + r\, n_d \tag{12}$$

which, for an integer r, equates the memory needed to store the given decomposition (left-hand side) and a decomposition in the same format, i.e. of an $n_1 \times \ldots \times n_d$-tensor, but with equal $d-1$ ranks $r, \ldots, r$ (right-hand side). In this sense, "effective" is understood with respect to memory. However, the notion of effective rank allows the exact evaluation of the complexity of some TT-structured operations, such as the matrix-vector multiplication and Hadamard product, and in a similar way estimates the complexity of other operations, e. g. the TT rank truncation.

2.2.2 Quantized Tensor Train representation

With the aim of further reduction of the complexity, the TT format can be applied to a "quantized" vector (matrix), which leads to the *Quantized Tensor Train* (QTT) format [32, 34, 33]. The idea of quantization consists in "folding" the vector (matrix) by introducing l_k "virtual" dimensions (levels) corresponding to the k-th original "physical" dimension [50], provided that the corresponding mode size n_k can be factorized as $n_k = n_{k,1} \cdot n_{k,2} \cdot \ldots \cdot n_{k,l_k}$ in terms of integral factors $n_{k,1}, \ldots, n_{k,l_k} \geq 2$, for $1 \leq k \leq d$. This corresponds to reshaping the k-th mode of size n_k into l_k modes of sizes $n_{k1}, \ldots, n_{k,l_k}$.

Under such a quantization applied to all dimensions, a d-dimensional $n_1 \times \ldots \times n_d$-vector indexed by $j_1 = \overline{j_{1,1}, \ldots, j_{1,l_1}}, \ldots, j_d = \overline{j_{d,1}, \ldots, j_{d,l_d}}$ is transformed into an $l_1 + \ldots + l_d$-dimensional $n_{1,1} \times \ldots \times n_{1,l_1} \times \ldots\ldots \times n_{d,1} \times \ldots \times n_{d,l_d}$-vector indexed by $j_{1,1}, \ldots, j_{1,l_1}, \ldots\ldots, j_{d,1}, \ldots, j_{d,l_d}$. A TT decomposition of the quantized vector is referred to as *QTT decomposition* of the original vector, the ranks of this TT decomposition are called *ranks of the QTT decomposition* of the original vector.

Example 2.3 (Proposition 1.1 in [34])**.** *To demonstrate how the quantization reduces complexity of structured data, let us consider the one-dimensional vector* $\boldsymbol{u}$ *whose entries are given by evaluation of the exponential with base* $q > 0$ *on the nonnegative integers* $\{0, 1, \ldots, 2^l - 1\}$*:* $\boldsymbol{u} = \left(1, q, \ldots, q^{2^l-1}\right)^\top$ *. Originally, there is only one dimension in this vector, and the elementwise representation requires storage of* 2^l *parameters since it does not exploit any structure in the data. However, if we use the quantization approach described above to split the single dimension into* l *virtual levels, the one-dimensional vector is transformed into* l*-dimensional one which exhibits a low-parametric structure. Indeed, in terms of the "virtual" indices it is a rank-one Kronecker product of* l *vectors with* 2 *components each:*

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ q^{2^{l-1}} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ q^{2^{l-2}} \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} 1 \\ q \end{pmatrix},$$

*which implies both rank-*1 *CP and QTT decompositions of* $\boldsymbol{u}$*. Other explicit low-rank examples can be found in [51, 52, 53, 54].*

If the natural ordering

$$\underbrace{j_{1,1}, \ldots, j_{1,l_1}}_{\text{1st dimension}}, \underbrace{j_{2,1}, \ldots, j_{2,l_2}}_{\text{2nd dimension}}, \ldots\ldots, \underbrace{j_{d,1}, \ldots, j_{d,l_d}}_{d\text{th dimension}} \tag{13}$$

of the "virtual" indices is used for representing the quantized vector in the TT format, then the ranks of the QTT decomposition can be enumerated as follows:

$$\underbrace{r_{1,1}, \ldots, r_{1,l_1-1}}_{\text{1st dimension}}, \hat{r}_1, \underbrace{r_{2,1}, \ldots, r_{2,l_2-1}}_{\text{2nd dimension}}, \hat{r}_2, \ldots\ldots, \hat{r}_{d-1}, \underbrace{r_{d,1}, \ldots, r_{d,l_d-1}}_{d\text{th dimension}},$$

where we emphasize $\hat{r}_1, \ldots, \hat{r}_{d-1}$ are the TT ranks of the original tensor, i.e. the ranks of the separation of "physical" dimensions. That is, the TT ranks of the tensor before quantization are conserved through the quantization process, until the reapproximation of the quantized tensor is concerned.

In this sense (10) and (11), with d being replaced with l, also present QTT representations of ranks $r_1, \ldots, r_{l-1}$ of a one-dimensional vector $\tilde{\boldsymbol{p}}$ and of a one-dimensional matrix $\tilde{\boldsymbol{A}}$ with entries

$\tilde{\boldsymbol{p}}_{\overline{j_1,\ldots,j_l}} = \boldsymbol{p}_{j_1,\ldots,j_l}$ and $\tilde{\boldsymbol{A}}_{\substack{\overline{i_1,\ldots,i_l}\\ \overline{j_1,\ldots,j_l}}} = \boldsymbol{A}_{\substack{i_1,\ldots,i_l\\ j_1,\ldots,j_l}}$ respectively. As a QTT decomposition is a TT decomposition of an appropriately quantized (and possibly, as we discuss in Section 2.2.5, transposed) tensor, the TT arithmetics referred to in Section 2.2.1, when applied to QTT decompositions, naturally provides the same basic operations in the QTT format.

Compared to the TT representation, the QTT format is able to resolve more structure in the data by splitting also the "virtual" dimensions introduced by the quantization.

Quantization is crucial for reducing the computational complexity further, as it allows the TT decomposition to seek and represent more structure in the data. In practice it appears the most efficient to use as fine quantization (i.e. with small n_{k,m_k}) as possible and in order to generate as many virtual modes as possible. As an example, when $n_k = 2^{l_k}$ for $1 \le k \le d$, one may consider the *ultimate quantization* with $n_{k,m_k} = 2$ for all m_k and k, so that $j_k = \overline{j_{k,1},\ldots,j_{k,l_k}} = \sum_{m_k=1}^{l_k} 2^{l_k-m_k} j_{k,m_k}$, where the indices $j_1,\ldots,j_l$ take the values 0 and 1.

The storage cost and complexity of basic QTT-structured operations are estimated from above through $d\,l\,r^\alpha$ with $\alpha \in \{2,3\}$, where $l \ge l_1,\ldots,l_d$ and r is an upper bound on all the QTT ranks of the decomposition in question. Note that this estimate may be, depending on r, logarithmic in n (also in $n^d = 2^{dl}$, which is an upper bound on the number of entries). The notion of an effective rank defined by (12) for TT decompositions applies verbatim to vectors and matrices represented in the QTT format.

2.2.3 The structure of the CME operator in the QTT format

In the following we consider the Finite State Projection of the CME, as described in Section 2.1, with $n_k = 2^{l_k}$ for $1 \le k \le d$ and assume that the PDF $\boldsymbol{p}$ of the truncated model and of the CME operator $\boldsymbol{A}$ from (3) are represented in the QTT format outlined in Section 2.2.2. We use the ultimate quantization, so that $n_{km} = 2$ for $1 \le m \le l_k$ and $1 \le k \le d$. In this section we mathematically establish rigorous upper bounds on the QTT ranks of $\boldsymbol{A}$ under certain assumptions on the propensity vectors $\boldsymbol{\omega}^s$, $1 \le s \le R$, defined by (8).

Theorem 2.4. *Consider the projected CME operator $\boldsymbol{A}$ defined by* (3)*. Assume that for every $s = 1,\ldots,R$ and $k = 1,\ldots,d$ the one-dimensional vector $\boldsymbol{\omega}_k^s$ from* (8)–(9) *is given in a QTT decomposition of ranks bounded by r_k^s; and that $\eta_k^s = 0$ implies $r_k^s = 1$. Then for $\boldsymbol{A}$ there exists a QTT decomposition of ranks*

$$q_1,\ldots,q_1,\hat{q}_1,q_2,\ldots,q_2,\hat{q}_2,\ldots,\ldots,\hat{q}_{d-1},q_d,\ldots,q_d$$

with $\hat{q}_k = R$ for $1 \le k \le d-1$ and

$$q_k = \sum_{\substack{s=1,\ldots,R:\\ \eta_k^s=0}} 2 \quad + \sum_{\substack{s=1,\ldots,R:\\ \eta_k^s\neq 0}} 3r_k^s$$

for $1 \le m_k \le l_k - 1$ and $1 \le k \le d$.

The proof is given in the supplement. □

A crude upper bound on the QTT ranks of the CME operator, following from Theorem 2.4 in terms of $r = \max_{s,k} r_k^s$, equals $3 \cdot R \cdot r$ and is still favorable, since it ensures the estimate $\mathcal{O}\left(d\,l R^2 r^2\right)$ for the number of parameters, i.e. the storage cost, where $l_1,\ldots,l_d \le l$. Note that if the kth factor ω_k^s of the s-th propensity function is a polynomial of degree p_k^s, then $\boldsymbol{\omega}_k^s$ (9) can be represented in the QTT format with ranks bounded by $r_k^s = p_k^s + 1$ uniformly in l_k, see [44, Corollary 13] and [51, Theorem 6]. In particular, this is the case when the reaction network is composed entirely of elementary reactions. Our numerical experiments show that the QTT ranks of propensity vectors corresponding to rational propensity functions are low as well, which results in low QTT ranks of the CME operator (see Section 2.3.3).

2.2.4 Transposed QTT representation

So far we have shown that the CME operator (3) under the FSP projection admits the low-parametric representation in the standard QTT format introduced in Sections 2.2.1–2.2.2. However, such a compressibility of the operator does not imply that the format is suitable for the efficient numerical solution of the CME. The following example demonstrates a simple example of non-axis-parallel features in the data, which cannot be represented in the format with low ranks. Our numerical experiments show that such features in the data may arise in systems with a conservation relationship between two or more chemical species resulting in a strong correlation in their copy numbers.

To illustrate this, let us consider the identity matrix and its vectorization:

$$\boldsymbol{A}_{\substack{i\\j}} = \boldsymbol{u}_{i,j} = \delta(i,j) \quad \text{for} \quad 1 \le i,j \le n, \tag{14}$$

where $n = 2^l$. The matrix $\boldsymbol{A}$, which is the only TT unfolding and the lth QTT unfolding of the vector $\boldsymbol{u}$, is of full rank. This implies that an exact representation of $\boldsymbol{u}$ in the formats described for vectors in Sections 2.2.1–2.2.2 will have at least one rank equal to 2^l and cannot represent $\boldsymbol{u}$ efficiently. However, the matricization $\boldsymbol{A}$ is perfectly separable:

$$\boldsymbol{A} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes l},$$

with a QTT matrix decomposition consisting of l cores V_k of size $1 \times 2 \times 2 \times 1$, given by

$$V_k(1,\cdot,\cdot,1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and with QTT ranks equal to $1,\dots,1$. In other words, the indices $i = \overline{i_1,\dots,i_l}$ and $j = \overline{j_1,\dots,j_l}$ may not be separable at all, while the mixed and re-ordered indices $\overline{i_1,j_1},\dots,\overline{i_l,j_l}$ are perfectly separable. The ordering of the multi-index should reflect the structure in the data to achieve an optimal compression.

The example above hints at a natural modification of the QTT decomposition. We represent in the TT format the quantized vector with virtual dimensions permuted so that the "virtual" indices corresponding to the same levels of quantization of different physical dimensions are adjacent; for example, for $l_1 = \ldots = l_d = l$ instead of (13) we use the ordering

$$\underbrace{j_{1,1},\dots,j_{d,1}}_{\text{1st level}},\underbrace{j_{1,2},\dots,j_{d,2}}_{\text{2nd level}},\cdots\cdots,\underbrace{j_{1,l},\dots,j_{d,l}}_{d\text{th level}}. \tag{15}$$

When $l_1,\dots,l_d$ are not equal, in order to obtain a similar to (15) transposed ordering of indices, we introduce void indices j_{k,m_k} with $n_{k,m_k} = 1$ for $l_k + 1 \le m_k \le \max_{1\le k'\le d} l_{k'}$, reorder all the "virtual" indices according to (15) and then drop the void ones. This modification of the QTT format, which we refer here to as *quantized-and-transposed Tensor Train*; shortly, *transposed QTT* or *QT3*. It was first applied to vectors in [55]; namely, to vectors of the form

$$\boldsymbol{u}_{j_1,\dots,j_d} = \begin{cases} 1, & \sum_{k=1}^d j_k \le 2^l, \\ 0, & \text{otherwise}, \end{cases}$$

where $\underline{j} = (j_1,\dots,j_d) \in \{1,\dots,2^l\}^d$. Such a vector may be considered as a discretization of the indicator function of the simplex $\{x \in \mathbb{R}^d_{\ge 0} : \|x\|_1 \le 1\}$. In [55], $\boldsymbol{u}$ was shown to have a QT3 decomposition of ranks bounded linearly in d uniformly in l. In the particular case of $d = 2$ such a bound follows from the result of [53] on the structure of Toeplitz matrices generated by tensor-structured vectors.

The index ordering (15) aims at the low-rank representation of such tensors, in which the physical dimensions are coupled on the corresponding virtual levels, i.e. *scales*, much more than different scales are within each single dimension. This is the case for the extreme example (14), where we end up with a rank-one decomposition if we choose to separate the scales first, and the physical dimensions, then. Despite such a difference in approximation properties, from the algorithmic point of view, QT3 is a minor modification of the standard, widely used form of the QTT format. We do not imply any particular ordering of indices by simply referring to QTT.

2.2.5 The structure of the CME operator in the transposed QTT format

Similarly to Theorem 2.4, we can bound the ranks of the CME operator in the transposed QTT format relying on the ordering (15) of "virtual" indices.

Theorem 2.5. *Consider the projected CME operator* $\boldsymbol{A}$ *defined by* (3)*. Assume that for every* $s = 1, \dots, R$ *and* $k = 1, \dots, d$ *the one-dimensional vector* $\boldsymbol{\omega}_k^s$ *from* (8)–(9) *is given in a QTT decomposition of ranks bounded by* r_k^s*; and that* $\eta_k^s = 0$ *implies* $r_k^s = 1$*. Then for* $\boldsymbol{A}$ *there exists a QT3 decomposition of ranks bounded by*

$$\sum_{s=1}^{R} \left(1 + \prod_{k \in \mathcal{K}^s} 2\right) \left(\prod_{k \in \mathcal{K}^s} r_k^s\right),$$

where $\mathcal{K}^s = \{k \in \mathbb{N} : 1 \le k \le d \text{ and } \eta_k^s \ne 0\}$.

The proof is given in the supplement. □

As Section 2.3.4 shows, the QT3 ranks of the CME operator may be significantly lower.

2.3 Numerical experiments

2.3.1 Common details

In the presentation of our numerical experiments, we use the following notation for the parameters of the DMRG solver: the required relative residual `RES` of the linear system, the maximum number `SWP` of its iterations ("sweeps"), the maximum number `RST` of GMRES restarts for the "local problem" of the DMRG optimization procedure, the maximum number `ITR` of such iterations before a restart, the maximum feasible rank `RMX`, the rank `KCK` of random components added to the solution to avoid stagnation. The DMRG iterations continue until either their number reaches `SWP` or the relative residual is less than or equal to `RES`. In every particular run all those parameters are the same for all time steps.

The fact that the DMRG solver, as any other tensor-structured solver available, converges only locally, requires the time steps to be rather small to allow for the corresponding linear systems being solved. For this reason, we have to use an equidistant mesh of mesh width h on $[h, T_1]$, where the transient processes are strong, but on $[T_1, T]$ can increase the mesh width geometrically with the grading factor $\sigma_2 = \frac{t_{m-1}}{t_m} = 1 - \frac{h}{T_1}$, which is only slightly less than 1. On $[0, h]$ we initialize our algorithm by $M_0 = 10$ steps graded geometrically with the factor $\sigma_0 = \frac{t_{m-1}}{t_m} = 0.5$. On all time intervals we use polynomial spaces of degree $\rho = 3$ to discretize (2) as described in Section 6.1, since the aforementioned limitation of the DMRG solver prevents us from using high polynomial degrees and enjoying the exponential convergence of the time discretization. For the bases in polynomial spaces corresponding to the time steps we take the orthonormal system of normalized Legendre polynomials.

At the mth time step, after having obtained $\boldsymbol{P}_m$ as an approximate solution of the corresponding linear system (20), we evaluate $\boldsymbol{p}_m^-$ and reapproximate it in the TT format with relative ℓ_2-accuracy `EPS` in order to drop excessive QTT components.

We compare the evaluated solution or its marginal to a reference data. By Δ_{ℓ_p} we denote the ℓ_p-norm of the discrepancy. Generally we start with the ℓ_2-norm, which can be easily computed even when the comparison is made only in the (Q)TT format and cannot be made in the

	Direct Approach		Proposed Approach					
run	solution	operator	solution		truncated solution		operator	
	Mem	Mem	Mem	ratio	Mem	ratio	Mem	ratio
d independent birth-death processes								
$d=1$	$4.10_{\mathbf{3}}$	$1.68_{\mathbf{7}}$	736	$1.80_{\mathbf{-1}}$	264	$6.45_{\mathbf{-2}}$	992	$5.91_{\mathbf{-5}}$
$d=2$	$1.68_{\mathbf{7}}$	$2.82_{\mathbf{14}}$	3858	$2.30_{\mathbf{-4}}$	528	$3.15_{\mathbf{-5}}$	2852	$1.01_{\mathbf{-11}}$
$d=3$	$6.87_{\mathbf{10}}$	$4.72_{\mathbf{21}}$	7742	$1.13_{\mathbf{-7}}$	898	$1.31_{\mathbf{-8}}$	4800	$1.02_{\mathbf{-18}}$
$d=4$	$2.81_{\mathbf{14}}$	$7.90_{\mathbf{28}}$	12176	$4.33_{\mathbf{-11}}$	1432	$5.09_{\mathbf{-12}}$	6748	$8.52_{\mathbf{-26}}$
$d=5$	$1.15_{\mathbf{18}}$	$1.32_{\mathbf{36}}$	16262	$1.41_{\mathbf{-14}}$	1946	$1.69_{\mathbf{-15}}$	11032	$8.30_{\mathbf{-33}}$
genetic toggle switch								
only	$3.36_{\mathbf{7}}$	$1.12_{\mathbf{15}}$	65264	$1.95_{\mathbf{-3}}$	–	–	10988	$9.76_{\mathbf{-12}}$
enzymatic futile cycle								
(A)	$4.19_{\mathbf{6}}$	$1.76_{\mathbf{13}}$	18396	$4.39_{\mathbf{-3}}$	8472	$2.02_{\mathbf{-3}}$	25848	$1.47_{\mathbf{-9}}$
(D)			360332	$8.59_{\mathbf{-2}}$	290144	$6.92_{\mathbf{-2}}$	5584	$3.17_{\mathbf{-10}}$

Table 1: Overview of the QTT compression of the storage needed for solutions (maximum throughout the time stepping) and CME operators. For details on "truncated solution" see Section 2.3.1. Solution Mem in the Direct Approach is the number of states taken into account in the FSP, which is equal to the number of entries, N, in the solution vector. For the CME operator, Mem is N^2, the number of entries in the matrix. In the Proposed QTT Approach, *ratio* indicates the memory storage compression ratio, i.e. the ratio of Mem in the Proposed QTT Approach to that in the Direct Approach. The exponents are given in boldface for the base 10.

full format (which is the case in Section 2.3.2 for $d \geq 3$). In some cases we compute also the discrepancy for $p = 1$ and the probability deficiency $\texttt{ERR}_\Sigma\,[\boldsymbol{p}_m^-\,] = |1 - \sum \boldsymbol{p}_m^-\,|$. The reference data is also obtained with a certain accuracy which cannot be reduced arbitrarily. Moreover, in some cases our solution appears to be more accurate, which accounts for using the term "discrepancy" instead of "error".

In the first and third examples we reapproximate the solution once more, but this time with with relative ℓ_2-accuracy $\alpha \cdot \frac{\Delta_{\ell_2}}{\|\boldsymbol{p}_m^-\|}$, where α is 0.05 and 0.01 respectively. Below we refer to this procedure as *truncation*, and the approximated vector, as *truncated solution*. The procedure ensures that the relative discrepancy in the ℓ_2-norm grows by the factor of $1 + \alpha$ at most and shows what QTT ranks allow for our numerical solution, obtained without using any reference data, to ensure *almost* the same discrepancy from the reference data (which is related to the accuracy of both the solution and reference data) as before truncation.

2.3.2 d Independent Birth-Death Processes

As a first example we consider a system composed of d chemical species with $\{X_1, \ldots, X_d\}$ a vector of random variables representing the species count of each. The dynamics of the random vector are governed by independent birth-death processes. For the k-th species, the corresponding reactions are given by

$$\emptyset \underset{d_k}{\overset{b_k}{\rightleftharpoons}} \mathrm{X}_k$$

where b_k is the spontaneous creation rate and d_k is the destruction rate for species X_k. The dynamics of any one chemical species of this system is independent of the dynamics of all others. Given the initial condition $X_k(0) = \xi_k$ for each k, the marginal distribution for any one species X_k at time t is given by:

$$p_k(x_k; t) = \mathcal{P}(x_k, \lambda_k(t)) \star_{x_k} \mathcal{M}(x_k, \xi_k, p^{(k)}(t)), \quad x_k \in \mathbb{Z}_{\geq 0}$$

where $\mathcal{P}(\cdot, \lambda_k(t))$ is the Poisson distribution with parameter $\lambda_k(t)$, $\star_{x_k}$ indicates the discrete convolution in variable x_k, $\mathcal{M}\big(x_k, \xi_k, p^{(k)}(t)\big)$ the multinomial distribution with parameter $p^{(k)}(t)$,

d	N	$\frac{\|\boldsymbol{A}\boldsymbol{p}_0\|_2}{\|\boldsymbol{p}_0\|_2}$	$\frac{\|\boldsymbol{A}\boldsymbol{p}_M^-\|_2}{\|\boldsymbol{p}_M^-\|_2}$	r_{eff}	Δ_{ℓ_2}	TIME
1	2^{12}	$1.4_{+\mathbf{3}}$	$1.0_{-\mathbf{3}}$	3.53	$1.9_{-\mathbf{5}}$	87
2	2^{24}	$2.4_{+\mathbf{3}}$	$1.4_{-\mathbf{3}}$	3.42	$2.3_{-\mathbf{5}}$	704
3	2^{36}	$3.5_{+\mathbf{3}}$	$1.8_{-\mathbf{3}}$	3.38	$3.5_{-\mathbf{5}}$	1548
4	2^{48}	$4.5_{+\mathbf{3}}$	$2.0_{-\mathbf{3}}$	3.37	$3.6_{-\mathbf{5}}$	2516
5	2^{60}	$5.5_{+\mathbf{3}}$	$2.3_{-\mathbf{3}}$	3.36	$3.5_{-\mathbf{5}}$	3544

Table 2: d independent birth-death processes: $r_{\text{eff}} = r_{\text{eff}}\left[\boldsymbol{p}_M^-\right]$, $\Delta_{\ell_2} = \Delta_{\ell_2}\left[\boldsymbol{p}_M^-\right]$, computational TIME in seconds; $r_{\max}\left[\boldsymbol{p}_M^-\right] = 6$ for all d. N is the number of states taken into account in the FSP. The exponents are given in boldface for the base 10.

and the parameters $p^{(k)}$ and λ_k evolve according to the reaction rate equations

$$\begin{gathered}\frac{\mathrm{d}}{\mathrm{d}t}\, p^{(k)}(t) = -d_k p^{(k)}(t)\,, \quad \frac{\mathrm{d}}{\mathrm{d}t}\,\lambda_k(t) = b_k - d_k\lambda_k(t), \\ p^{(k)}(0) = 1, \qquad\qquad \lambda_k(t) = 0.\end{gathered}$$

See [16, Theorem 1] for details. Since $X_1, \ldots, X_k$ are mutually independent, the full joint PDF at time t, $\boldsymbol{p}(t)$, is the product of the marginals:

$$\boldsymbol{p}(t) = \prod_{k=1}^{d} p_k(t)$$

that is, this system has an explicit formula for the solution regardless of the number of chemical species involved. We can, therefore, evaluate the accuracy and observe the complexity scaling of the *hp*-DG-QTT solver as the number of chemical species increases.

For numerical simulations we assume $b_k = 1000$ and $d_k = 1$ for $1 \le k \le d$ and consider the FSP with $l_k = 12$. We solve the corresponding projected CME for $d = 1, 2, 3, 4, 5$ to check that in all these cases the *hp*-DG-QTT method using the ordering (13) without transposition is capable of revealing the same low-rank QTT structure of the solution. For the CME operator we have $r_{\max}[\boldsymbol{A}] \le 8$ up to accuracy $5 \cdot 10^{-15}$.

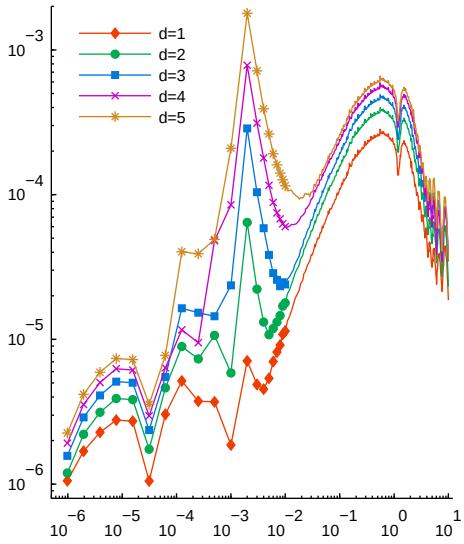
For a zero initial value we run the time stepping till $T = 10$ with $T_1 = 10^{-1}$ and $h = 10^{-3}$, which takes $M = 569$ steps overall. The settings of the DMRG solver are: RES $= 2 \cdot 10^{-6}$, SWP $= 2$, RMX $= 20$, ITR $= 100$, RST $= 1$, KCK $= 1$. The evaluation accuracy is EPS $= 10^{-8}$. The results, which are presented in Figure 2 and Table 2, show that the same low-rank structure of the solution is adaptively reconstructed by the algorithm for all d considered. The transient phase causes the growth of QTT ranks, because at certain steps of every sweep the DMRG solver merges virtual dimensions corresponding to different species and attempts to adapt the rank separating them. As a consequence, the transient phase is passed with overestimated ranks, but at larger times the QTT structure of the numerical solution is the same.

2.3.3 Toggle Switch

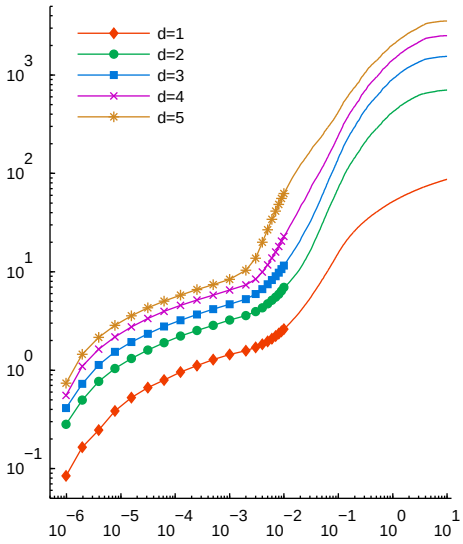
The next example models a synthetic gene-regulatory circuit designed to produce bistability over a wide range of parameter values [56]. The network is composed of two repressors and two constitutive promoters arranged in a feedback loop so that each promoter is inhibited by the repressor transcribed by the opposing promoter 3. This mutually inhibitory arrangement gives rise to the robust bistable behavior of the network. If the concentration of one repressor is high, this lowers the production rate of the other repressor, keeping its concentration low. This allows a high rate of production of the original repressor, thereby stabilizing its high concentration.

A stochastic model of the toggle switch was considered in [57] and consists of the following four reactions:

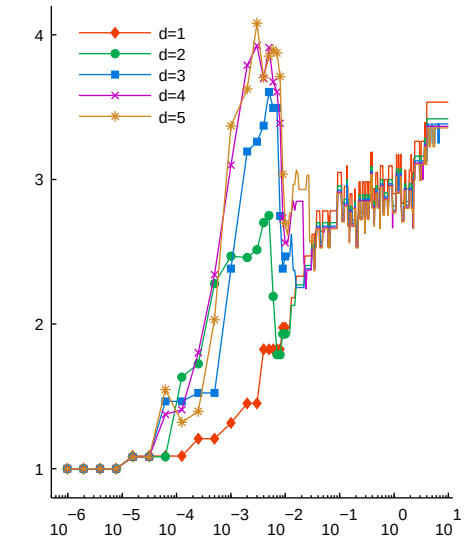
$$\emptyset \xrightarrow{\frac{\alpha_1}{1+V^\beta}} \mathrm{U}, \quad \emptyset \xrightarrow{\frac{\alpha_2}{1+U^\gamma}} \mathrm{V}$$



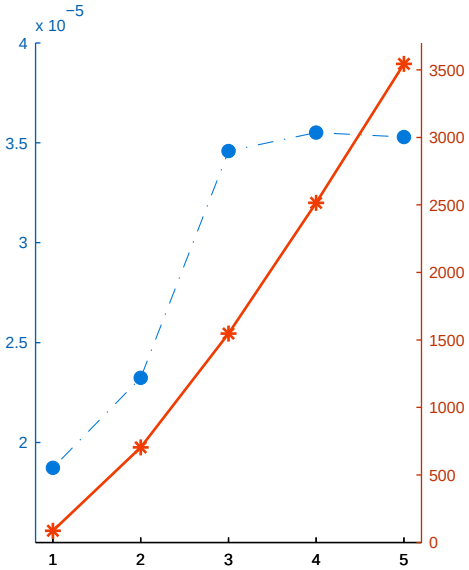
(a) $\Delta_{\ell_2}\left[\boldsymbol{p}_m^-\right] / \left\|\boldsymbol{p}_M^-\right\|$ (after truncation) vs. t_m



(b) Cumulative computation time vs. t_m



(c) Effective QTT rank $r_{\text{eff}}\left[\boldsymbol{p}_m^-\right]$ (after truncation) vs. t_m



(d) Relative discrepancy $\Delta_{\ell_2}\left[\boldsymbol{p}_M^-\right] / \left\|\boldsymbol{p}_M^-\right\|$ (blue) and total computation time (red) vs. d

Figure 2: d independent birth-death processes. Computation time is given in seconds; $r_{\max}\left[\boldsymbol{p}_M^-\right] = 6$ for all d. Markers are omitted for $t_m > 10^{-2}$ in (a)–(c)

.

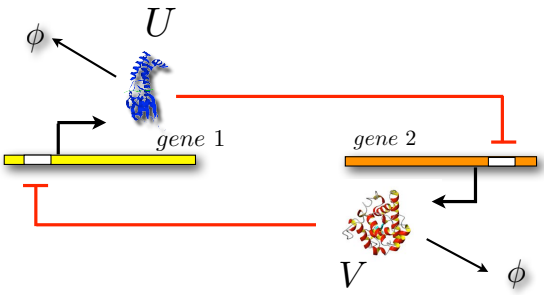


Figure 3: Toggle Switch consisting of double negative feedback loop.

$$\mathrm{U} \xrightarrow{\delta_1} \emptyset, \qquad \mathrm{V} \xrightarrow{\delta_2} \emptyset$$

where U and V represent the two repressors. Denote the species counts of each by U and V, respectively. The stochastic model admits a bimodal stationary distribution over a wide range of values of the rate constants. We consider the set of parameters from [57] which were selected to test the efficiency of using available numerical algorithms to calculate matrix exponentials to solve low dimensional FSP approximations of the CME. We then scaled the parameters so that a larger set of states would be required to guarantee an FSP truncation with low approximation error. While a different set of parameters were considered in [58, 20], which required a larger FSP truncation, this choice of values renders the system symmetric under interchange of the roles of U and V. This situation is less biologically relevant than what we consider here.

For this numerical example we assume $\alpha_1 = 5000$, $\alpha_2 = 1600$, $\beta = 2.5$, $\gamma = 1.5$, $\delta_1 = \delta_2 = 1$. We consider the FSP with $l_U = 13$, $l_V = 12$, which allows to take into account 2^{25} states. The initial value is zero. We use the ordering (13) without transposition. For the CME operator we have $r_{\max}[\boldsymbol{A}] = 14$ and $r_{\text{eff}}[\boldsymbol{A}] = 10.89$ up to accuracy 10^{-14}. The settings of the DMRG solver are: $\texttt{RES} = 10^{-6}$, $\texttt{SWP} = 3$, $\texttt{RMX} = 200$, $\texttt{ITR} = 100$, $\texttt{RST} = 2$, $\texttt{KCK} = 2$. The evaluation accuracy is $\texttt{EPS} = 10^{-8}$. We model the dynamics of the CME till $T = 100$ with $T_1 = 10$ and $h = 0.03$, which takes $M = 1111$ steps overall.

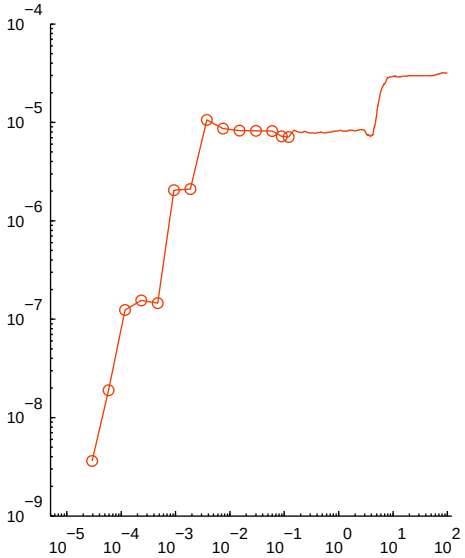
The results are presented in Figure 4. At the terminal time T we have $\texttt{ERR}_\Sigma\left[\boldsymbol{p}_M^-\right] = 3.17 \cdot 10^{-5}$. The validation with the PDF based on 816 million Monte Carlo simulations (every 1000 draws taking on average over 360 seconds, adding up to an overall CPU time over $3 \cdot 10^8$ seconds) yields $\Delta_{\ell_1}\left[\boldsymbol{p}_M^-\right] = 8.34 \cdot 10^{-4}$, and for the 2- and Chebyshev norms we have $\Delta_{\ell_2}\left[\boldsymbol{p}_M^-\right] / \left\|\boldsymbol{p}_M^-\right\|_2 = 6.62 \cdot 10^{-4}$ and $\Delta_{\ell_\infty}\left[\boldsymbol{p}_M^-\right] = 5.50 \cdot 10^{-6}$. As for the ranks, $r_{\text{eff}}\left[\boldsymbol{p}_M^-\right] = 8.74$ and $r_{\max}\left[\boldsymbol{p}_M^-\right] = 13$. Figure 4c shows that after $t \approx 20$ the norm of the time derivative stagnates at approximately 10^{-5} determined by the accuracy parameters chosen, and the following time steps require negligible computational effort. At the same time, as we see in Figure 4b, all QTT ranks stabilize under 15, but the transient phase preceding that moment involves far higher ranks. Figure 5a presents a snapshot of the distribution.

2.3.4 Enzymatic Futile Cycle

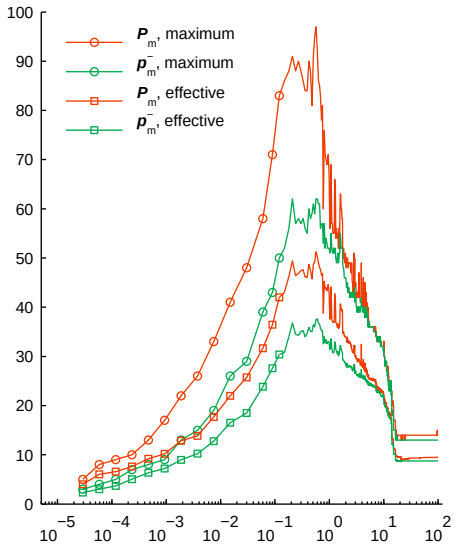
Futile cycles are composed of two metabolic or signaling pathways that work in opposite directions meaning that the products of one pathway are the precursors for the other and vice versa 6. This biochemical network structure results in no net production of molecules and often results only in the dissipation of energy as heat [59]. Nevertheless, there is an abundance of known pathways that use this motif and it is thought to provide a highly tunable control mechanism with potentially high sensitivity [59, 60].

[60] introduced a stochastic version of the model with just the essential network components required to model the dynamics. The stochastic model consists of six chemical species and six reactions:

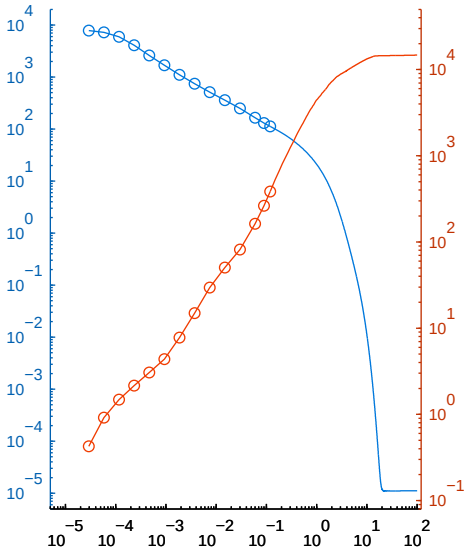
$$\mathrm{X} + \mathrm{E}_+^{\mathrm{f}} \underset{k_{+2}}{\overset{k_{+1}}{\rightleftharpoons}} \mathrm{E}_+^{\mathrm{b}}, \quad \mathrm{X}^* + \mathrm{E}_-^{\mathrm{f}} \underset{k_{-2}}{\overset{k_{-1}}{\rightleftharpoons}} \mathrm{E}_-^{\mathrm{b}},$$



(a) Probability deficiency $\texttt{ERR}_\Sigma\left[\boldsymbol{p}_m^-\right]$

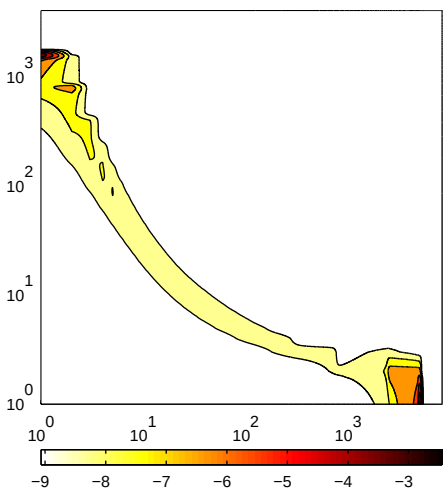


(b) Maximum and effective QTT ranks

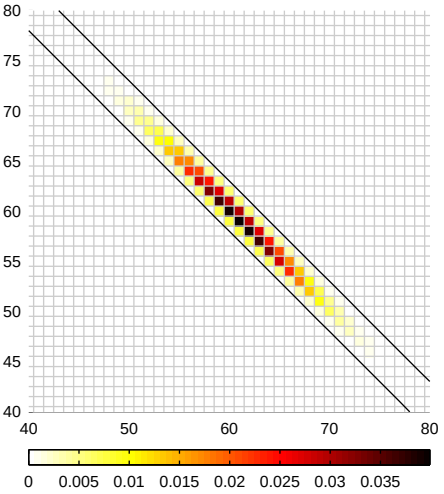


(c) Relative norm $\frac{\|\boldsymbol{A}\boldsymbol{p}_m^-\|_2}{\|\boldsymbol{p}_m^-\|_2}$ of the derivative (blue) and cumulative computation time (red, sec.)

Figure 4: Genetic toggle switch. The values are given vs. t_m. Markers are omitted for $t_m > 10^{-1}$.



(a) Genetic toggle switch. The PDF for $m = 350$, $t_m \approx 10.18$, U (hor.) vs. V (vert.). As the process evolves, the probability mass becomes concentrated in two distinct regions. Contour coloring is logarithmically scaled with base 10.



(b) Enzymatic futile cycle. The marginal PDF for $m = 20$, $t_m = 5 \cdot 10^{-3}$, X (vert.) vs. X^* (hor.). Black lines delimit the states reachable from the initial condition. The transposed QTT format automatically exploits this sparsity pattern *of the full PDF* for compression without special input from the user.

Figure 5: Snapshots of solutions.

$$\mathrm{E}_+^{\mathrm{b}} \xrightarrow{k_{+3}} \mathrm{E}_+^{\mathrm{f}} + \mathrm{X}^*, \quad \mathrm{E}_-^{\mathrm{b}} \xrightarrow{k_{-3}} \mathrm{E}_-^{\mathrm{f}} + \mathrm{X},$$

$\{\mathrm{X}, \mathrm{X}^*\}$ represent the forward substrate and product, $\{\mathrm{E}_+, \mathrm{E}_-\}$ denote the forward and reverse enzymes, respectively. Note that this system is closed meaning that particles are neither created nor destroyed. We denote the random variables representing the molecule count of each species with italics.

For the particular set of initial conditions considered in [60] the number of states that are reachable is large enough to render a direct numerical solution of the CME impractical. The authors instead used the Gillespie Direct SSA to generate a large number of sample paths to estimate the distribution. The authors also applied a diffusion approximation to their model which resulted in a SDE which produced qualitatively similar dynamics. To the authors' knowledge, no attempt has been made so far towards the direct numerical solution of the CME for this system.

At time t, let $X^{\mathrm{T}}(t)$ denote the total amount of both free and bound substrate, and $E_+^{\mathrm{T}}(t)$ and $E_-^{\mathrm{T}}(t)$ the total forward and reverse enzymes, respectively. We observe the following conservation relations:

$$E_+^{\mathrm{f}}(t) + E_+^{\mathrm{b}}(t) = E_+^{\mathrm{T}}(t) = E_+^{\mathrm{T}}(0)$$

$$E_-^{\mathrm{f}}(t) + E_-^{\mathrm{b}}(t) = E_-^{\mathrm{T}}(t) = E_-^{\mathrm{T}}(0)$$

$$X(t) + X^*(t) + E_+^{\mathrm{b}}(t) + E_-^{\mathrm{b}}(t) = X^{\mathrm{T}}(t) = X^{\mathrm{T}}(0)$$

Using the above, one can establish an upper and lower bound relating the species count of $\mathrm{X}(t)$ to $\mathrm{X}^*(t)$ that depends only on the total initial amount of substrate and the total initial amount of enzymes in the system

$$X^{\mathrm{T}}(0) - X^*(t) \geq X(t) \geq X^{\mathrm{T}}(0) - X^*(t) - \left(E_+^{\mathrm{T}}(0) + E_-^{\mathrm{T}}(0)\right).$$

Assuming that the initial quantity of enzymes $E_+^{\mathrm{T}}(0) + E_-^{\mathrm{T}}(0)$ is small, for a given copy number of $X^*(t)$, $X(t)$ may take at most $E_+^{\mathrm{T}}(0) + E_-^{\mathrm{T}}(0)$ different values. Since $X^{\mathrm{T}}(t)$ is a conserved quantity, this means that $X(t)$ and $X^*(t)$ will be strongly anti-correlated. Under these circumstances, we

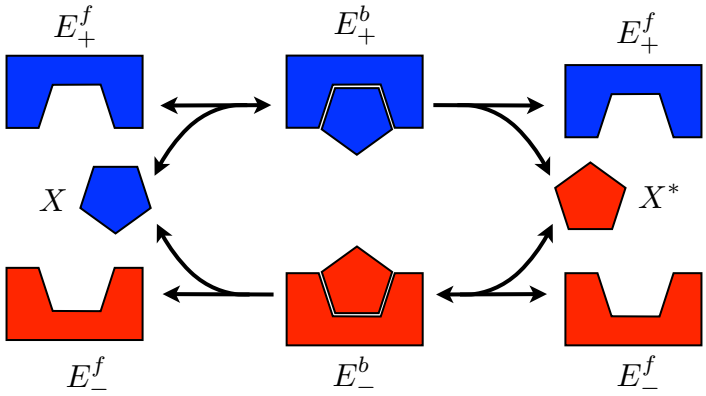


Figure 6: Enzymatic Futile Cycle.

find in our numerical experiments that the transposed QTT format is better suited than the standard QTT to efficiently represent the corresponding PDF.

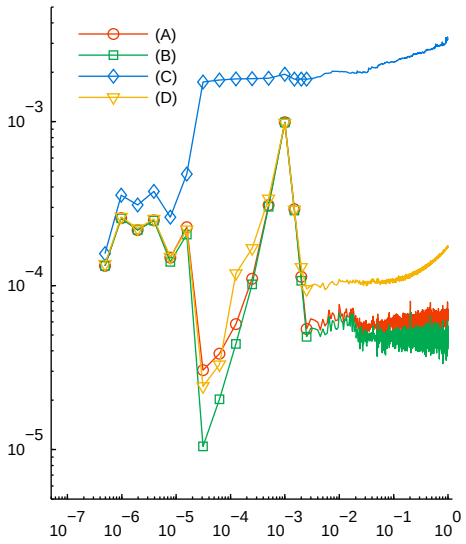
Following [60], we consider $k_{+1} = 40$, $k_{+2} = 10^4$, $k_{+3} = 10^4$, $k_{-1} = 200$, $k_{-2} = 100$, $k_{-3} = 5000$. For initial value we take $E_\pm^f = 2, E_\pm^b = 0, X = 30, X^* = 90$. We consider the FSP projection with $l_{E_\pm^{b,f}} = 2$ and $l_X = l_{X^*} = 7$, i.e. with 2^{22} states. We present 4 runs: (A), (B) and (C) use the transposed QTT format, and (D), the standard QTT. Theorems 2.5 and 2.4 bound the exact QTT ranks of the CME operator by 216 and 21 respectively, and numerically for accuracy 10^{-14} we have $r_{\max}[\boldsymbol{A}] = 38$, $r_{\text{eff}}[\boldsymbol{A}] = 17.93$ in (A)–(C) and $r_{\max}[\boldsymbol{A}] = 11$, $r_{\text{eff}}[\boldsymbol{A}] = 8.30$ in (D). We model the dynamics of the CME till $T = 1$ with $T_1 = 0.3$ and $h = 5 \cdot 10^{-4}$, which takes $M = 1332$ steps overall. For (A) and (D), which differ in the format, we keep the same accuracy parameters: $\texttt{RES} = 10^{-6}$ and $\texttt{EPS} = 10^{-8}$. On the other hand, (B) and (C) use the same format as (A), but different accuracy parameters. In (B) they are $\texttt{RES} = 10^{-8}$ and $\texttt{EPS} = 10^{-10}$; in (C), $\texttt{RES} = 10^{-4}$ and $\texttt{EPS} = 10^{-6}$. As a result, (B) and (C) provide, respectively, a more accurate and a cruder solution as compared to (A).

This experiment shows, in particular, that lower ranks of the operator do not necessarily lead to lower ranks of the solution, and that the transposed QTT format actually ensures smaller ranks of the solution in this example. We set $\texttt{RMX} = 200$ in (A)–(C) and $\texttt{RMX} = 400$ in (D) and observe that $\max_{0 \le t_m \le 0.1} r_{\max}[\boldsymbol{P}_m]$ reaches 51 for (A) and 359 for (D). Other parameters of the DMRG solver are the same for all 4 runs: $\texttt{SWP} = 5$, $\texttt{ITR} = 50$, $\texttt{RST} = 2$, $\texttt{KCK} = 2$.

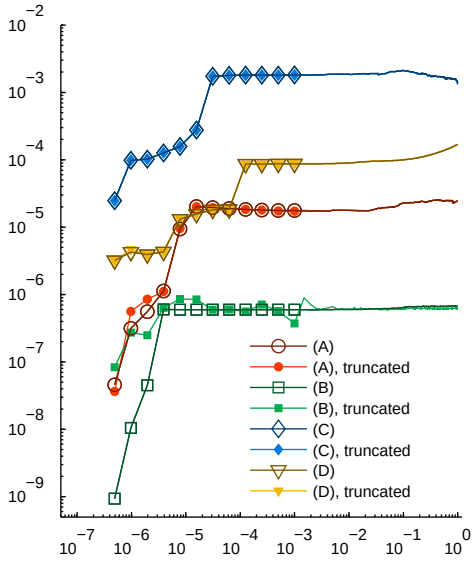
For every m, we validate our solution $\boldsymbol{p}_m^-$ by comparing its marginal distribution $\sum_{E_\pm^{\text{b,f}}} \boldsymbol{p}_m^-$ to that based on $18.6 \cdot 10^9$ Monte Carlo simulations (every 10000 draws taking at least 110 seconds, amounting to an overall CPU time over $2 \cdot 10^8$ seconds). The discrepancy $\Delta_{\ell_p} = \Delta_{\ell_p}\left[\sum_{E_\pm^{\text{b,f}}} \boldsymbol{p}_m^-\right]$ in the marginal distribution with respect to X and X^* is reported for $p = 1$ in Figure 7a and Table 3. With $p = 2$ we use it for the discrepancy-based truncation, which, as Figure 7b shows, does not affect the probability deficiency significantly.

Figure 7a shows that the refined run (B) yields the smallest discrepancy, which suggests that the reference distribution is sufficiently accurate to allow for the discrepancy to represent the actual error in the results of (A), (B) and (C). As we can see from Figure 7d, in all 4 runs the time derivative stagnates after $t \approx 0.1$, at lower levels for more accurate runs. Let us note that at that stage in (A)–(C) it exhibits relatively strong oscillations compared to (D), which happens due to different effect of the addition of random components in the DMRG solver in the presence and absence of the transposition. On the other hand, compared to (A), the run (D) yields a less accurate solution and reaches $t = 0.1$ almost 9 times later, the accuracy settings being the same in these two runs. In all, the transposition appears to make the QTT format far more efficient in this experiment, and we expect it to be even more so in larger systems of such type.

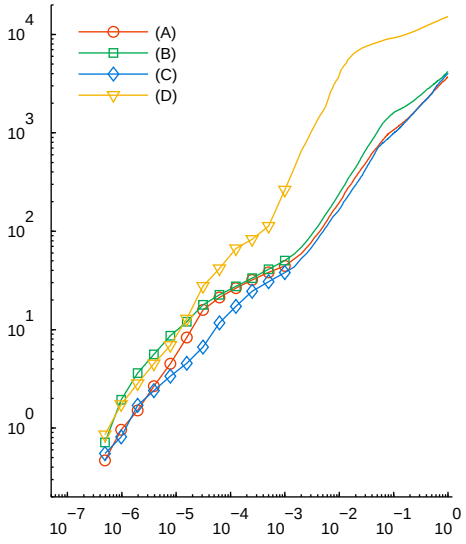
The results are given in Figures 7 and 8 and in Table 3. Figure 5b presents a snapshot of the marginal distribution.



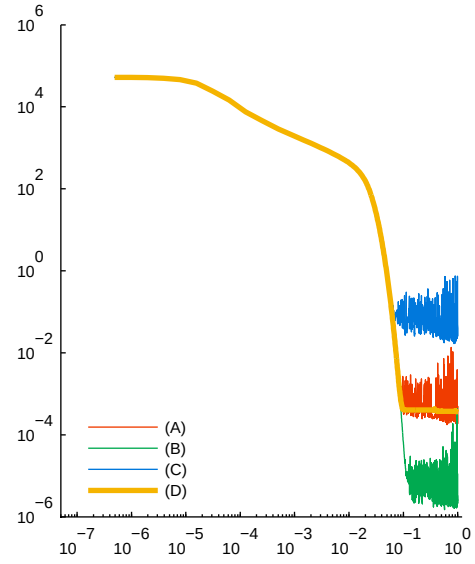
(a) Discrepancy Δ_{ℓ_1} (before truncation) from the marginal PDF based on Monte Carlo simulations



(b) Probability deficiency $\mathtt{ERR}_{\Sigma}\left[\boldsymbol{p}_m^-\right]$

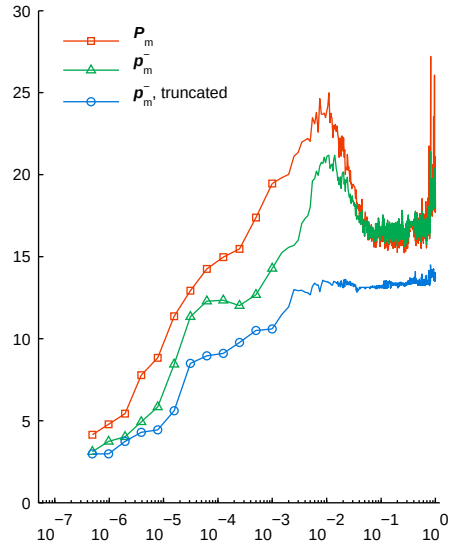


(c) Cumulative computation time (sec.)

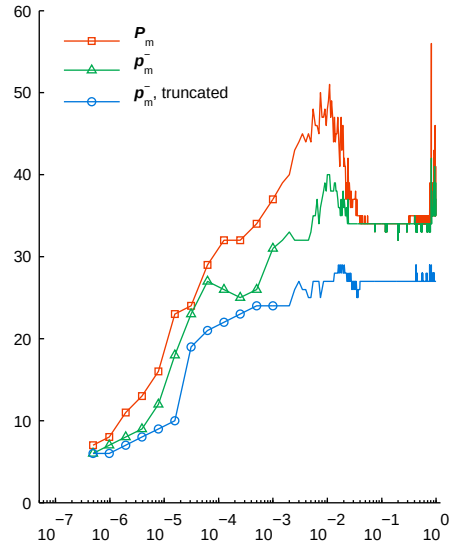


(d) Relative norm $\frac{\|\boldsymbol{A}\boldsymbol{p}_m^-\|_2}{\|\boldsymbol{p}_m^-\|_2}$ of the derivative

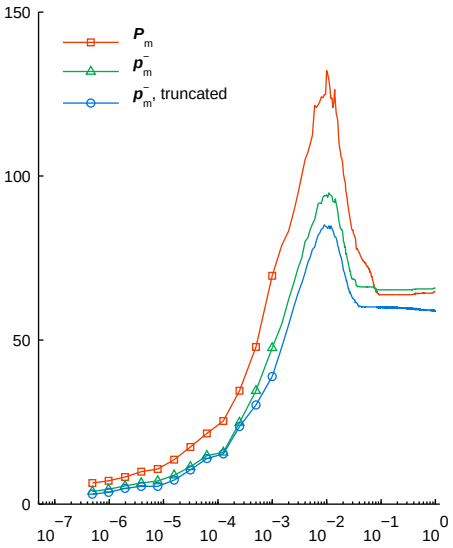
Figure 7: Enzymatic futile cycle. The values are given vs. t_m. Markers are omitted for $t_m \geq 2 \cdot 10^{-3}$ in 7b–7c



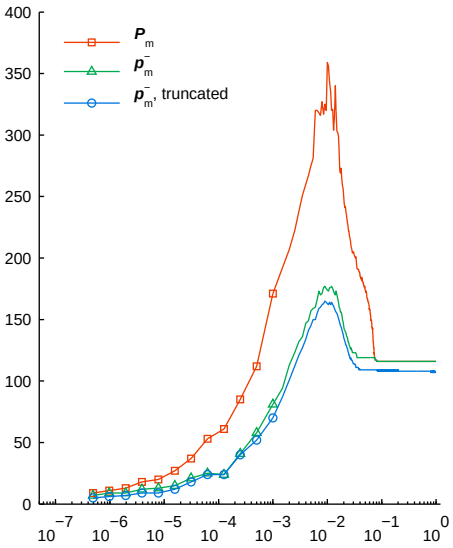
(a) Effective QTT ranks r_{eff} for (A)



(b) Maximum QTT ranks $r_{\max}$ for (A)



(c) Effective QTT ranks r_{eff} for (D)



(d) Maximum QTT ranks $r_{\max}$ for (D)

Figure 8: Enzymatic futile cycle. QTT ranks of the solution. The values are given vs. t_m. Markers are omitted for $t_m \geq 2 \cdot 10^{-3}$

run	$\frac{\|\boldsymbol{A}\boldsymbol{p}_m^-\|_2}{\|\boldsymbol{p}_m^-\|_2}$	r_{eff}	$r_{\max}$	Δ_{ℓ_1}	$\texttt{ERR}_\Sigma$	TIME
$m = 210$, $t_m = 0.1$						
(A)	$3.5_{\,\mathbf{-4}}$	13.17	27	$5.7_{\,\mathbf{-5}}$	$2.3_{\,\mathbf{-5}}$	$1.07_{\,\mathbf{3}}$
(B)	$6.5_{\,\mathbf{-5}}$	12.14	25	$4.6_{\,\mathbf{-5}}$	$6.1_{\,\mathbf{-7}}$	$1.60_{\,\mathbf{3}}$
(C)	$1.3_{\,\mathbf{-1}}$	12.16	24	$2.3_{\,\mathbf{-3}}$	$2.1_{\,\mathbf{-3}}$	$9.87_{\,\mathbf{2}}$
(D)	$4.1_{\,\mathbf{-4}}$	60.06	109	$1.1_{\,\mathbf{-4}}$	$1.0_{\,\mathbf{-4}}$	$9.23_{\,\mathbf{3}}$
$m = M = 1332$, $t_m = T = 1$						
(A)	$1.8_{\,\mathbf{-4}}$	13.66	27	$7.2_{\,\mathbf{-5}}$	$2.5_{\,\mathbf{-5}}$	$3.70_{\,\mathbf{3}}$
(B)	$1.1_{\,\mathbf{-5}}$	12.06	25	$5.7_{\,\mathbf{-5}}$	$6.2_{\,\mathbf{-7}}$	$4.21_{\,\mathbf{3}}$
(C)	$2.5_{\,\mathbf{-2}}$	12.85	24	$3.3_{\,\mathbf{-3}}$	$1.3_{\,\mathbf{-3}}$	$4.03_{\,\mathbf{3}}$
(D)	$3.7_{\,\mathbf{-4}}$	58.97	107	$1.7_{\,\mathbf{-4}}$	$1.7_{\,\mathbf{-4}}$	$1.52_{\,\mathbf{4}}$

Table 3: Enzymatic futile cycle: $r_{\text{eff}} = r_{\text{eff}}[\boldsymbol{p}_m^-]$, $r_{\max} = r_{\max}[\boldsymbol{p}_m^-]$, $\Delta_{\ell_1} = \Delta_{\ell_1}\left[\sum_{E_\pm^{\text{b,f}}} \boldsymbol{p}_m^-\right]$, $\texttt{ERR}_\Sigma = \texttt{ERR}_\Sigma[\boldsymbol{p}_m^-]$ are given for the truncated solution $\boldsymbol{p}_m^-$; computational TIME is given in seconds; $\frac{\|\boldsymbol{A}\boldsymbol{p}_0\|_2}{\|\boldsymbol{p}_0\|_2} = 5.2 \cdot 10^4$. The exponents are given in boldface for the base 10.

3 Methods

To solve the initial value problem for (2), we exploit the *hp*-DG-QTT algorithm proposed in [35], implemented in MATLAB. It uses an implicit, exponentially convergent spectral time discretization of discontinuous Galerkin type (see Sections 6.1 and 6.2). Discretization of the resulting, time-discrete CME in "species space" is done in the QTT format. Our realization of this implementation relies on the public domain *TT Toolbox* which provides basic TT-structured operations and solvers for linear systems in the QTT format. The TT toolbox is publicly available at `http://spring.inm.ras.ru/osel` and `http://github.com/oseledets/TT-Toolbox`; to be consistent, we use the GitHub version of July 12, 2012 in all examples below. We run the *hp*-DG-QTT solver in MATLAB 7.12.0.635 (R2011a) on a laptop with a 2.7 GHz dual-core processor and 4 GB RAM, and report the computational time in seconds.

For the solution of the large, linear systems in the QTT and QT3 formats in each time step, we use the DMRG optimization solver, proposed for the TT format in [61] and available as the function `dmrg_solve3` of the TT Toolbox. The DMRG solver still lacks a rigorous theoretical foundation. In [62] a closely related *Alternating Least Squares (ALS)* approach was mathematically analyzed and shown to converge at least locally. However, in the high order implicit time discretizations of the CME considered in this paper, the DMRG solver proved to be highly efficient. More on the mathematical ideas behind the ALS and DMRG optimization in the TT format can be found in [63]. We remark that while there is currently no estimate of the convergence rate for the DMRG algorithm, in our numerical experiments reported below we found the solver to be highly efficient. The DMRG solver, under certain restrictions on the time step, manages to find a parsimonious QTT formatted solution of the linear system (up to a specified tolerance). Moreover, the solver in effect automatically adapts the both the QTT rank as well as the QTT "basis" of the solution at every time step guaranteeing that it is sufficiently rich in order to capture the principal dynamics of interest.

In the first numerical example the solution is symmetric and exactly rank-one separable, which allows us to use the standard MATLAB solver `ode15s` in the sparse format to obtain the univariate factor of a reference solution. In other examples we used SPSens beta 3.4 massively parallel package for the stochastic simulation of chemical networks (`http://sourceforge.net/projects/spsens/`) [64], to construct reference PDFs. Those computations were carried out on up to 1500 cores of Brutus, the central high-performance cluster of ETH Zürich (`http://www.clusterwiki.ethz.ch/brutus/Brutus_wiki`).

4 Conclusion

We presented a novel, "ab-initio" computational methodology for the direct numerical solution of the CME. The methodology exploits the time-analytic nature of solutions to the CME and the low-rank, tensor structure of the CME operator by combining an *hp*-timestepping method that is order and step size adaptive, unconditionally stable and exponentially convergent with respect to the number of time discretization parameters, with novel, tensor-formatted linear algebra techniques for the numerical realization of the method. In particular, after an initial projection on a (sufficiently rich) finite state, the so-called Quantized, Tensor-Train (QTT for short) formatted numerical linear algebra affords dynamic adaptation of the state-space size, as well as of the principal components, or basis elements of the numerical representation of solution vectors in the numerical simulation of the time evolution of the CME solution. The approach is, therefore, superior to fixed basis approaches, even when used with adaptivity, such as those reported in [14, 20, 65, 19].

As we mention above, the performance of the approach proposed essentially relies on the efficiency of the solver of TT-structured linear system. In particular, a globally (or "less strictly locally") convergent iterative solver would allow us to take larger time steps and to exploit the exponential convergence of the *hp*-DG time discretization. We believe that while the presently reported numerical results which were obtained with the DMRG solver are quite encouraging, ongoing research on TT-structured linear system solvers holds the promise for a substantial efficiency increase of the present methodology. We only mention a family of alternating minimal energy methods which was announced very recently in [66].

We also mention that, of course, the choice of tensor format and, possibly, index ordering, has an essential impact on the performance of the approach. The computational experiments reported in Section 2.3.4 of the present paper show that even a straightforward permutation of "virtual" indices produced by quantization may allow to exploit additional structure in the data and the QTT formatted CME solution and, therefore, may improve the performance of the QTT-structured approach dramatically. We point out that the TT format can be considered as a special case of Tensor Network States: TT formatted tensor are tensor networks in which the tensor network has the form of a simple, rooted tree. A general discussion of tensor networks and their use in numerical simulations for quantum spin systems can be found in [67, 68]. As for the numerical solution of the CME, particular real-life problems might require more sophisticated tensor networks to be used to efficiently approximate reachable states of the systems in question. The mathematical investigation of the relative merits and drawbacks of tensor formats for particular applications is currently undergoing rather active development; we mention only the recent monograph [37] and the references there.

We finally mention that recently, and independently, TT formatted linear algebra methods for the CME were proposed in [69]; a low order time stepping, and no transposition of tensor trains was used in this reference. The CME examples presented in [69] also included a toggle switch. Unfortunately, the paper does not contain mathematical convergence results, and no attempt is made to quantify, even for the numerical examples considered, the numerical errors for the numerical solutions obtained, for example by comparison with benchmark numerical results obtained with other simulation methods. The comparisons in the present paper with state-of-the-art, massively parallel stochastic simulation packages, however, allow on the one hand, validation of accuracy of the QTT-based solutions obtained here and, on the other hand, also evidence the dramatic increase in efficiency afforded by the new deterministic approach: Monte Carlo simulations on 1500 cores of a high-performance cluster were matched in accuracy and outperformed in the wall-clock time by a MATLAB implementation running on a notebook.

5 Author contributions

Conception of the approach: VK MK MN CS. Implementation of the *hp*-DG-QTT approach and of the transposed QTT format: VK. Selected models for and assisted with the exper-

iments: MN. Designed the experiments, analyzed the data: VK MN. Performed the experiments: VK. Wrote the paper: VK MK MN CS.

References

[1] *M. Elowitz, A. Levine, E. Siggia, P. Swain.* Stochastic Gene Expression in a Single Cell // *Nature*. 2002. V. 297, No. 5584. P. 1183–1186. 2

[2] *H. H. McAdams, A. Arkin.* Stochastic mechanisms in gene expression // *PNAS*. 1997. V. 94, No. 3. P. 814–819. DOI: 10.1073/pnas.94.3.814. 2

[3] *D. T. Gillespie.* A general method for numerically simulating the stochastic time evolution of coupled chemical reactions // *Journal of Computational Physics*. 1976. V. 22, No. 4. P. 403–434. DOI: 10.1016/0021-9991(76)90041-3. http://www.sciencedirect.com/science/article/B6WHY-4DD1NC9-CP/2/43ade5f11fb949602b3a2abdbbb29f0e. 2

[4] *M. A. Gibson, J. Bruck.* Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels // *The Journal of Physical Chemistry A*. 2000. V. 104, No. 9. P. 1876-1889. DOI: 10.1021/jp993732q. http://pubs.acs.org/doi/pdf/10.1021/jp993732q. http://pubs.acs.org/doi/abs/10.1021/jp993732q. 2

[5] *D. T. Gillespie.* Approximate accelerated stochastic simulation of chemically reacting systems // *The Journal of Chemical Physics*. 2001. V. 115, No. 4. P. 1716-1733. DOI: 10.1063/1.1378322. http://link.aip.org/link/?JCP/115/1716/1. 2

[6] *N. G. van Kampen.* Stochastic Processes in Physics and Chemistry. — Amsterdam and New York: North-Holland, 1992. 2, 4

[7] *J. a. P. Hespanha, A. Singh.* Stochastic Models for Chemically Reacting Systems Using Polynomial Stochastic Hybrid Systems // *Int. J. on Robust Control*, Special Issue on Control at Small Scales: Issue 1. 2005, Sep. V. 15. P. 669–689. 2

[8] *C. A. Gomez-Uribe, G. C. Verghese.* Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations // *The Journal of Chemical Physics*. 2007. V. 126, No. 2. P. 024109. DOI: 10.1063/1.2408422. http://link.aip.org/link/?JCP/126/024109/1. 2

[9] *D. T. Gillespie.* The chemical Langevin equation // *The Journal of Chemical Physics*. 2000. V. 113, No. 1. P. 297-306. DOI: 10.1063/1.481811. http://link.aip.org/link/?JCP/113/297/1. 2

[10] *S. N. Ethier, T. G. Kurtz.* Markov Processes: Characterization and Convergence. — New York: Wiley-Interscience, 2005. 2

[11] *J. Puchalka, A. M. Kierzek.* Bridging the Gap between Stochastic and Deterministic Regimes in the Kinetic Simulations of the Biochemical Reaction Networks // *Biophysical Journal*. 2004. V. 86, No. 3. P. 1357–1372. DOI: 10.1016/S0006-3495(04)74207-1. http://www.sciencedirect.com/science/article/pii/S0006349504742071. 2

[12] *E. L. Haseltine, J. B. Rawlings.* Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics // *The Journal of Chemical Physics*. 2002. V. 117, No. 15. P. 6959-6969. DOI: 10.1063/1.1505860. http://link.aip.org/link/?JCP/117/6959/1. 2

[13] *H. Salis, Y. Kaznessis.* Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions // *The Journal of Chemical Physics*. 2005. V. 122, No. 5. P. 054103. DOI: 10.1063/1.1835951. http://link.aip.org/link/?JCP/122/054103/1. 2

[14] *A. Hellander, P. Lötstedt.* Hybrid method for the chemical master equation // *Journal of Computational Physics*. 2007. V. 227, No. 1. P. 100–122. DOI: DOI: 10.1016/j.jcp.2007.07.020. http://www.sciencedirect.com/science/article/pii/S0021999107003221. 2, 22

[15] *T. Jahnke.* On Reduced Models for the Chemical Master Equation // *Multiscale Modeling and Simulation*. 2011. V. 9. P. 1646. 2

[16] *T. Jahnke, W. Huisinga.* Solving the chemical master equation for monomolecular reaction systems analytically // *Journal of mathematical biology*. 2007. V. 54, No. 1. P. 1–26. 2, 4, 13

[17] *B. Munsky, M. Khammash.* The finite state projection algorithm for the solution of the chemical master equation // *The Journal of Chemical Physics*. 2006. V. 124, No. 4. P. 044104. DOI: 10.1063/1.2145882. http://link.aip.org/link/?JCP/124/044104/1. 2, 4, 5

[18] *T. Henzinger, M. Mateescu, V. Wolf.* Sliding Window Abstraction for Infinite Markov Chains // Computer Aided Verification / Ed. by A. Bouajjani, O. Maler. — Springer Berlin / Heidelberg, 2009. — V. 5643 of *Lecture Notes in Computer Science*. — P. 337-352. — 10.1007/978-3-642-02658-4-27. http://dx.doi.org/10.1007/978-3-642-02658-4_27. 2

[19] *S. Engblom.* Spectral approximation of solutions to the chemical master equation // *Journal of Computational and Applied Mathematics*. 2009. V. 229, No. 1. P. 208–221. DOI: 10.1016/j.cam.2008.10.029. http://www.sciencedirect.com/science/article/pii/S0377042708005578. 2, 22

[20] *P. Deuflhard, W. Huisinga, T. Jahnke, M. Wulkow.* Adaptive Discrete Galerkin Methods Applied to the Chemical Master Equation // *Sci. Comput.* 2008. V. 30, No. 6. P. 2990–3011. 2, 15, 22

[21] *M. Hegland, C. Burden, L. Santoso et al.* A solver for the stochastic master equation applied to gene regulatory networks // *Journal of computational and applied mathematics*. 2007. V. 205, No. 2. P. 708–724. 2

[22] *T. Jahnke, T. Udrescu.* Solving chemical master equations by adaptive wavelet compression // *Journal of Computational Physics*. 2010. V. 229, No. 16. P. 5724–5741. 2

[23] *R. Bellman.* Adaptive Control Processes: A Guided Tour. — Princeton, NJ: Princeton University Press, 1961. 3

[24] *F. L. Hitchcock.* The expression of a tensor or a polyadic as a sum of products // *Journal of Mathematics and Physics*. 1926, November. V. 6, No. 1. P. 164–189. 3

[25] *J. D. Caroll, J. J. Chang.* Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition // *Psychometrika*. 1970. V. 35. P. 283–319. DOI: 10.1007/BF02310791. http://dx.doi.org/10.1007/BF02310791. 3

[26] *M. Hegland, J. Garcke.* On the numerical solution of the chemical master equation with sums of rank one tensors // *ANZIAM Journal*. 2011. V. 52, No. 0. http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/3895. 3, 4

[27] *V. de Silva, L.-H. Lim.* Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem // *SIAM Journal on Matrix Analysis and Applications*. 2008. V. 30, No. 3. P. 1084–1127. DOI: 10.1137/06066518X. http://epubs.siam.org/sima/resource/1/sjmael/v30/i3/p1084_s1. 3

[28] *J. Håstad*. Tensor rank is NP-complete // *Journal of Algorithms*. 1990. V. 11, No. 4. P. 644–654. DOI: 10.1016/0196-6774(90)90014-6. http://www.sciencedirect.com/science/article/pii/0196677490900146. 3

[29] *C. Hillar, L.-H. Lim*. Most tensor problems are NP hard // *arXiv*. 2009. V. abs/0911.1393. http://arxiv.org/abs/0911.1393. 3

[30] *T. Jahnke, W. Huisinga*. A dynamical low-rank approach to the chemical master equation // *Bulletin of mathematical biology*. 2008. V. 70, No. 8. P. 2283–2302. 3

[31] *I. V. Oseledets, E. E. Tyrtyshnikov*. Breaking the curse of dimensionality, or how to use SVD in many dimensions // *SIAM Journal on Scientific Computing*. 2009, October. V. 31, No. 5. P. 3744–3759. DOI: 10.1137/090748330. http://epubs.siam.org/sisc/resource/1/sjoce3/v31/i5/p3744_s1. 3, 6

[32] *I. Oseledets*. Approximation of matrices with logarithmic number of parameters // *Doklady Mathematics*. 2009. V. 80. P. 653–654. DOI: 10.1134/S1064562409050056. http://dx.doi.org/10.1134/S1064562409050056. 3, 8

[33] *I. V. Oseledets*. Approximation of $2^d \times 2^d$ matrices using tensor decomposition // *SIAM Journal on Matrix Analysis and Applications*. 2010. V. 31, No. 4. P. 2130–2145. DOI: 10.1137/090757861. http://link.aip.org/link/?SML/31/2130/1. 3, 8

[34] *B. N. Khoromskij*. $\mathcal{O}(d \log N)$-Quantics Approximation of N-d Tensors in High-Dimensional Numerical Modeling // *Constructive Approximation*. 2011. V. 34, No. 2. P. 257–280. DOI: 10.1007/s00365-011-9131-1, 10.1007/s00365-011-9131-1. http://www.springerlink.com/content/06n7q85q14528454/. 3, 8

[35] *V. Kazeev, O. Reichmann, C. Schwab*. *hp*-DG-QTT solution of high-dimensional degenerate diffusion equations: Research Report 11: Seminar for Applied Mathematics, ETH Zürich, 2012. http://www.sam.math.ethz.ch/reports/2012/11. 3, 21, 29

[36] *T. G. Kolda, B. W. Bader*. Tensor Decompositions and Applications // *SIAM Review*. 2009, September. V. 51, No. 3. P. 455–500. DOI: 10.1.1.153.2059. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.153.2059&rep=rep1&type=pdf. 5

[37] *W. Hackbusch*. Tensor Spaces and Numerical Tensor Calculus. — Springer, 2012. — V. 42 of *Springer Series in Computational Mathematics*. http://www.springerlink.com/content/l62t86. 5, 7, 22

[38] *I. V. Oseledets*. Tensor Train decomposition // *SIAM Journal on Scientific Computing*. 2011. V. 33, No. 5. P. 2295–2317. DOI: 10.1137/090752286. http://dx.doi.org/10.1137/090752286. 6, 7, 32

[39] *S. R. White*. Density-matrix algorithms for quantum renormalization groups // *Phys. Rev. B*. 1993, October. V. 48, No. 14. P. 10345–10356. DOI: 10.1103/PhysRevB.48.10345. http://link.aps.org/doi/10.1103/PhysRevB.48.10345. 7

[40] *F. Verstraete, D. Porras, J. I. Cirac*. Density Matrix Renormalization Group and Periodic Boundary Conditions: A Quantum Information Perspective // *Phys. Rev. Lett.* 2004, November. V. 93, No. 22. P. 227205. DOI: 10.1103/PhysRevLett.93.227205. http://link.aps.org/doi/10.1103/PhysRevLett.93.227205. 7

[41] *G. Vidal*. Efficient Classical Simulation of Slightly Entangled Quantum Computations // *Phys. Rev. Lett.* 2003, October. V. 91, No. 14. P. 147902. DOI: 10.1103/PhysRevLett.91.147902. http://link.aps.org/doi/10.1103/PhysRevLett.91.147902. 7

[42] *W. Hackbusch, S. Kühn.* A New Scheme for the Tensor Representation // *Journal of Fourier Analysis and Applications*. 2009. V. 15, No. 5. P. 706–722. DOI: 10.1007/s00041-009-9094-9, 10.1007/s00041-009-9094-9. http://www.springerlink.com/content/t3747nk47m368g44. 7

[43] *L. Grasedyck.* Hierarchical Singular Value Decomposition of Tensors // *SIAM Journal on Matrix Analysis and Applications*. 2010. V. 31, No. 4. P. 2029–2054. DOI: 10.1137/090764189. http://link.aip.org/link/?SML/31/2029/1. 7

[44] *L. Grasedyck.* Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization: Preprint 308: Institut für Geometrie und Praktische Mathematik, RWTH Aachen, 2010, April. http://www.igpm.rwth-aachen.de/Download/reports/pdf/IGPM308_k.pdf. 7, 9

[45] *J. Ballani, L. Grasedyck.* A projection method to solve linear systems in tensor format // *Numerical Linear Algebra with Applications*. 2012. DOI: 10.1002/nla.1818. http://dx.doi.org/10.1002/nla.1818. 7

[46] *D. Kressner, C. Tobler.* Preconditioned low-rank methods for high-dimensional elliptic PDE eigenvalue problems // *Computational Methods in Applied Mathematics*. 2011. V. 11, No. 3. P. 363–381. http://cmam.info/index.php?do=issues/art&vol=11&num=3&art=323. 7

[47] *S. V. Dolgov, B. N. Khoromskij, I. V. Oseledets.* Fast solution of multi-dimensional parabolic problems in the TT/QTT-format with initial application to the Fokker-Planck equation (submitted to SISC): Preprint 80: Max-Planck-Institut für Mathematik in den Naturwissenschaften, 2011. http://www.mis.mpg.de/publications/preprints/2011/prepr2011-80.html. 7

[48] *D. Kressner, C. Tobler.* Low-rank tensor Krylov subspace methods for parametrized linear systems: Research Report 16: Seminar for Applied Mathematics, ETH Zürich, 2010. http://www.sam.math.ethz.ch/reports/2010/16. 7

[49] *B. N. Khoromskij, C. Schwab.* Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs // *SIAM Journal on Scientific Computing*. 2011. V. 33, No. 1. P. 364–385. http://epubs.siam.org/sisc/resource/1/sjoce3/v33/i1/p364_s1. 7

[50] *E. E. Tyrtyshnikov.* Tensor approximations of matrices generated by asymptotically smooth functions // *Sbornik: Mathematics*. 2003. V. 194, No. 5. P. 941–954. DOI: 10.1070/SM2003v194n06ABEH000747. http://iopscience.iop.org/1064-5616/194/6/A09. 8

[51] *I. V. Oseledets.* Constructive representation of functions in tensor formats // *Constructive Approximation*. 2013. No. 37. P. 1–18. http://link.springer.com/article/10.1007/s00365-012-9175-x. 8, 9

[52] *V. A. Kazeev, B. N. Khoromskij.* Low-rank explicit QTT representation of the Laplace operator and its inverse // *SIAM Journal on Matrix Analysis and Applications*. 2012. V. 33, No. 3. P. 742–758. DOI: 10.1137/100820479. http://epubs.siam.org/doi/abs/10.1137/100820479. 8, 29, 31

[53] *V. A. Kazeev, B. N. Khoromskij, E. E. Tyrtyshnikov.* Multiulevel Toeplitz matrices generated by tensor-structured vectors and convolution with logarithmic complexity: Preprint 36: Max-Planck-Institut für Mathematik in den Naturwissenschaften, 2011. http://www.mis.mpg.de/publications/preprints/2011/prepr2011-36.html. 8, 10, 29, 35

[54] *V. Kazeev, O. Reichmann, C. Schwab.* Low-rank tensor structure of linear diffusion operators in the TT and QTT formats // *Linear Algebra and its Applications*. 2013. DOI: 10.1016/j.laa.2013.01.009. 8, 29

[55] *I. V. Oseledets*. QTT decomposition of the characteristic function of a simplex: Personal communication: 2010, September. 10

[56] *T. S. Gardner, C. R. Cantor, J. J. Collins*. Construction of a genetic toggle switch in Escherichia coli // *Nature*. 2000. V. 403, No. 6767. P. 339-342. DOI: http://dx.doi.org/10.1038/35002131. http://www.nature.com/nature/journal/v403/n6767/suppinfo/403339a0_S1.html. 13

[57] *B. Munsky, M. Khammash*. The Finite State Projection Approach for the Analysis of Stochastic Noise in Gene Networks // *Automatic Control, IEEE Transactions on*. 2008, jan. V. 53, No. Special Issue. P. 201 -214. DOI: 10.1109/TAC.2007.911361. 13, 15

[58] *P. Sjöberg, P. Lötstedt, J. Elf*. Fokker–Planck approximation of the master equation in molecular biology // *Computing and Visualization in Science*. 2009. V. 12. P. 37-50. 10.1007/s00791-006-0045-6. http://dx.doi.org/10.1007/s00791-006-0045-6. 15

[59] *J. Schwender, J. Ohlrogge, Y. Shachar-Hill*. Understanding flux in plant metabolic networks // *Current Opinion in Plant Biology*. 2004. V. 7, No. 3. P. 309–317. DOI: 10.1016/j.pbi.2004.03.016. http://www.sciencedirect.com/science/article/pii/S1369526604000482. 15

[60] *M. Samoilov, S. Plyasunov, A. P. Arkin*. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations // *Proceedings of the National Academy of Sciences of the United States of America*. 2005. V. 102, No. 7. P. 2310-2315. DOI: 10.1073/pnas.0406841102. http://www.pnas.org/content/102/7/2310.full.pdf+html. http://www.pnas.org/content/102/7/2310.abstract. 15, 17, 18

[61] *S. V. Dolgov, I. V. Oseledets*. Solution of linear systems and matrix inversion in the TT-format (submitted to SISC): Preprint 19: Max-Planck-Institut für Mathematik in den Naturwissenschaften, 2011. http://www.mis.mpg.de/publications/preprints/2011/prepr2011-19.html. 21

[62] *T. Rohwedder, A. Uschmajew*. Local convergence of alternating schemes for optimization of convex problems in the TT format: Preprint 112: DFG-Schwerpunktprogramm 1324, 2012, August. http://www.dfg-spp1324.de/download/preprints/preprint112.pdf. 21

[63] *S. Holtz, T. Rohwedder, R. Schneider*. The alternating linear scheme for tensor optimization in the Tensor Train format // *SIAM Journal on Scientific Computing*. 2012. V. 34, No. 2. P. A683-A713. DOI: 10.1137/100818893. http://epubs.siam.org/doi/abs/10.1137/100818893. 21

[64] *P. W. Sheppard, M. Rathinam, M. Khammash*. SPSens: a software package for stochastic parameter sensitivity analysis of biochemical reaction networks // *Bioinformatics*. 2013. V. 29, No. 1. P. 140-142. DOI: 10.1093/bioinformatics/bts642. http://bioinformatics.oxfordjournals.org/content/29/1/140.full.pdf+html. http://bioinformatics.oxfordjournals.org/content/29/1/140.abstract. 21

[65] *T. Jahnke*. An adaptive wavelet method for the chemical master equation // *SIAM Journal on Scientific Computing*. 2010. V. 31. P. 4373. 22

[66] *S. V. Dolgov, D. V. Savostyanov*. Alternating minimal energy methods for linear systems in higher dimensions. Part I: SPD systems: arXiv preprint 1301.6068: 2013, January. http://arxiv.org/abs/1301.6068. 22

[67] *F. Verstraete, J. I. Cirac, V. Murg*. Matrix Product States, Projected Entangled Pair States, and variational renormalization group methods for quantum spin systems: arXiv preprint 0907.2796: 2009, July. http://arxiv.org/abs/0907.2796. 22

[68] *J. I. Cirac, F. Verstraete*. Renormalization and tensor product states in spin chains and lattices // *Journal of Physics A: Mathematical and Theoretical*. 2009. V. 42, No. 50. P. 504004. http://stacks.iop.org/1751-8121/42/i=50/a=504004. 22

[69] *S. V. Dolgov, B. N. Khoromskij*. Tensor-product approach to global time-space-parametric discretization of chemical master equation: Preprint 68: Max-Planck-Institut für Mathematik in den Naturwissenschaften, 2012, November. http://www.mis.mpg.de/publications/preprints/2012/prepr2012-68.html. 22

[70] *D. Schötzau, C. Schwab*. An hp a priori error analysis of the DG time-stepping method for initial value problems // *Calcolo*. 2000. V. 37. P. 207–232. DOI: 10.1007/s100920070002. http://dx.doi.org/10.1007/s100920070002. 29, 30

[71] *W. D. Launey, J. Seberry*. The strong Kronecker product // *Journal of Combinatorial Theory, Series A*. 1994. V. 66, No. 2. P. 192–213. DOI: 10.1016/0097-3165(94)90062-0. http://www.sciencedirect.com/science/article/pii/0097316594900620. 31

6 Supplementary material

First, we outline the *hp*-DG time stepping in Section 6.1 and discuss the tensor structure of resulting linear systems in Section 6.2. In Section 6.3 we revisit the notions of *core matrices* and *strong Kronecker product* according to the papers [52, 53, 54]. We use the notation introduced there to present in Section 6.4 some basic operations in the TT format. Finally, in Section 6.5 we provide proofs of Theorems 2.4, 2.5 and 6.4 and, therefore, for all assertions on QTT ranks made in the present paper. We note that the Theorems 2.4 and 2.5 are new mathematical results, and Lemma 6.11, on which the latter is based, may have applications well beyond the numerical solution of the CME.

6.1 *hp*-DG discretization of the CME

Let $X = \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_d} \sim \mathbb{R}^N$, where $N = n_1 \cdot \ldots \cdot n_d$, and consider the Cauchy problem for an autonomous ODE on a time interval $J = (0, T)$ with an operator $\boldsymbol{A} : X \to X$ and an initial value $\boldsymbol{p}_0 \in X$: find a continuously differentiable function $\boldsymbol{p} : \overline{J} \to X$ such that

$$\left\{ \begin{array}{rcl} \dot{\boldsymbol{p}}(t) & = & \boldsymbol{A} \cdot \boldsymbol{p}(t) \quad \text{for} \quad t \in \overline{J}, \\ \boldsymbol{p}(0) & = & \boldsymbol{p}_0 \, . \end{array} \right. \tag{16}$$

The corresponding Cauchy problem for (5) is a particular example of (16). The solution to (16) is given theoretically by

$$\boldsymbol{p}(t) = \exp\left(t\boldsymbol{A}\right) \cdot \boldsymbol{p}_0 \, , \quad t \in \overline{J}, \tag{17}$$

but the straightforward numerical evaluation of the matrix exponential involved is a very challenging task due to the "curse of dimensionality". Instead, we use the QTT-structured *Discontinuous Galerkin (DG)* time-stepping scheme, proposed in [35], to solve (16). For abstract, linear and autonomous initial value problems such as (16), which admits a unique solution which is an analytic function of time t and which takes values in the high-dimensional state space $\mathbb{R}^N$, the discontinuous-Galerkin (DG for short) time discretization was suggested and analyzed in detail earlier in [70]. In the presentation of the *hp*-DG-QTT algorithm for (16) below we rely on the latter paper in the presentation of the DG part, and on the former one, in the presentation of aspects related to the QTT structure of CME operator and of the tensors arising after the DG time semidiscretization.

To present the DG semidiscretization, we denote by $\mathcal{P}^{\rho}\left(I, X\right)$ the space of polynomials defined on a finite interval I, of degree at most ρ at most and with coefficients from the abstract space X.

Definition 6.1. *Let $\mathcal{M} = \{J_m\}_{m=1}^{M}$ be a partition of the time interval J into subintervals $J_m = (t_{m-1}, t_m)$, $1 \le m \le M$, and $\underline{\rho} \in (\mathbb{N} \cup \{0\})^M$. Consider the space*

$$\mathcal{P}^{\underline{\rho}}\left(\mathcal{M}, X\right) = \{\boldsymbol{p} : J \to X : \ \boldsymbol{p}|_{J_m} \in \mathcal{P}^{\rho_m}\left(J_m, X\right) \quad \textit{for} \quad 1 \le m \le M\}$$

of functions, which are polynomials of degree ρ_m at most on J_m for all m. Let $\boldsymbol{p}_m^+ = \lim_{t \downarrow t_m} \boldsymbol{p}(t)$ and $\boldsymbol{p}_m^- = \lim_{t \uparrow t_m} \boldsymbol{p}(t)$ and for all feasible m and for all $\boldsymbol{p} \in \mathcal{P}^{\underline{\rho}}\left(\mathcal{M}, X\right)$.

Then the Discontinuous Galerkin FEM formulation of (16)*, corresponding to the partition $\mathcal{M}$ and the vector of polynomial degrees $\underline{\rho}$, reads: find $\boldsymbol{p} \in \mathcal{P}^{\underline{\rho}}\left(\mathcal{M}, X\right)$ such that*

$$\sum_{m=1}^{M} \int_{J_m} \langle \dot{\boldsymbol{p}} - \boldsymbol{A}\boldsymbol{p}, \, \boldsymbol{q} \rangle \, \mathrm{d}t + \sum_{m=1}^{M} \left\langle \boldsymbol{p}_{m-1}^+ - \boldsymbol{p}_{m-1}^- \, , \, \boldsymbol{q}_{m-1}^+ \right\rangle = 0 \tag{18}$$

for all $\boldsymbol{q} \in \mathcal{P}^{\underline{\rho}}\left(\mathcal{M}, X\right)$, where $\boldsymbol{p}_0^-$ stands for the initial value $\boldsymbol{p}_0$.

Equation (18) can be understood as a time-stepping method: if $\boldsymbol{p}|_{J_m} \in \mathcal{P}^{\rho_m}\left(J_m, X\right)$ are known for $1 \le m \le \widehat{m} - 1$, then $\boldsymbol{p}|_{J_{\widehat{m}}} \in \mathcal{P}^{\rho_{\widehat{m}}}\left(J_{\widehat{m}}, X\right)$ can be found as the solution to

$$\int_{J_{\widehat{m}}} \langle \dot{\boldsymbol{p}} - \boldsymbol{A}\boldsymbol{p}, \, \boldsymbol{q} \rangle \, \mathrm{d}t + \left\langle \boldsymbol{p}_{\widehat{m}-1}^+ - \boldsymbol{p}_{\widehat{m}-1}^- \, , \, \boldsymbol{q}_{\widehat{m}-1}^+ \right\rangle = 0. \tag{19}$$

For $1 \le m \le M$ let $\{\phi_j\}_{j=0}^{\rho_m}$ be a basis in $\mathcal{P}^{\rho_m}((-1,1),X)$, then the corresponding temporal shape functions on J_m are $\phi_j \circ F_m^{-1}$, $0 \le j \le \rho_m$, where the affine map $F_m\colon (-1,1) \to J_m$ is defined by $t = F_m(\tau) = \frac{1}{2}(t_m+t_{m-1})+\frac{1}{2}(t_m-t_{m-1})\tau$ for $\tau \in (-1,1)$. If $\boldsymbol{p}|_{J_m} = \sum_{j=0}^{\rho_m} \left(\phi_j \circ F_m^{-1}\right) \cdot \boldsymbol{P}_{mj}$, where $\boldsymbol{P}_m \in X^{\rho_m+1}$, then (19) yields the following linear system on the coefficients:

$$(\boldsymbol{C}_m \otimes \mathbb{I} - \boldsymbol{G}_m \otimes \boldsymbol{A}) \cdot \boldsymbol{P}_m = \boldsymbol{\phi}_{m-1} \otimes \boldsymbol{p}_{m-1}^{-}, \tag{20}$$

where $(\boldsymbol{C}_m)_{\substack{i\\j}} = \int_{-1}^{1} \phi_j'(\tau)\,\phi_i(\tau)\,\mathrm{d}\tau + \phi_j(-1)\,\phi_i(-1)$ and $(\boldsymbol{G}_m)_{\substack{i\\j}} = \int_{-1}^{1} \phi_j(\tau)\,\phi_i(\tau)\mathrm{d}\tau$ for $0 \le i,j \le \rho_m$, while $\left(\boldsymbol{\phi}_{m-1}\right)_i = \phi_i(-1)$ for $0 \le i \le \rho_m$.

Let us denote $|\mathcal{M}| = \max_{1\le m\le M}(t_m - t_{m-1})$. A fixed point argument (valid even for nonlinear evolution equations with Lipschitz nonlinearity) was used in [70] to prove the following result.

Proposition 6.2 (Theorem 2.6 in [70])**.** *Assume that $\|\boldsymbol{A}\|_2 \cdot |\mathcal{M}| < 1$. Then there exists a unique solution to the linear tensor equations* (18) *which result from the DG time-semidiscretization of the CME.*

This existence result was complemented in [70] by a convergence rate estimate for the DG solutions.

Proposition 6.3. *Let $\hat{\boldsymbol{p}}$ and $\boldsymbol{p}$ be solutions of* (16) *and* (18) *respectively. Then*

$$\sup_{t\in\overline{J}} \|\boldsymbol{p}(t) - \hat{\boldsymbol{p}}(t)\|_2 \le C(\|\boldsymbol{A}\|_2, T) \cdot \widetilde{C}(\underline{\rho}) \cdot \max_{1\le m\le M} \left[(c|\mathcal{M}|)^{\rho_m+1} \cdot \rho_m^{-\rho_m-\frac{1}{4}} \cdot \exp \rho_m \right]$$

holds with a positive constant $c > 0$, where $|\underline{\rho}| = \max_{1\le m\le M} \rho_m$ and $\widetilde{C}(\underline{\rho}) = \log^{\frac{1}{2}} \max\left\{2, |\underline{\rho}|\right\}$.

The proof follows from Theorem 3.12 in [70] in the analytic case, and from Stirling's formula. □

The *hp*-DG time discretization allows, on the one hand, to resolve fast transients in the evolution by the (usual) time-step adaptation and, on the other hand, affords *order adaptation* for time-analytic solutions such as matrix exponentials of the CME operator. In particular, due to the time-analyticity of the solution, exponential rates of convergence in $\underline{\rho}$ are achieved, as can be seen from Proposition 6.3: for $\underline{\rho} = (\rho, \ldots, \rho)$ the error bound of Proposition 6.3 can be recast as

$$\sup_{t\in\overline{J}} \|\boldsymbol{p}(t) - \hat{\boldsymbol{p}}(t)\|_2 \le C \exp(-b\rho)$$

with constants $C, b > 0$ asymptotically independent of ρ, see [70, Theorem 3.18]. This implies that a prescribed level of accuracy ε can be reached with $\rho M = \mathcal{O}\left(\log \varepsilon^{-1}\right)$ temporal degrees of freedom.

6.2 QTT structure of time-step linear systems

Let us assume that the matrix $\boldsymbol{A}$ is represented in the QTT or QT3 format in terms of $\tilde{d}$ cores. In particular, if $n_k = 2^{l_k}$ for $1 \le k \le d$, then $\tilde{d} = l_1 + \ldots + l_d$ for the ultimate quantization. The system (20) is of order $\rho_m \times n_1 \times \ldots \times n_d$, where the first dimension accounts for the coefficients $\boldsymbol{P}_m$ of the solution $\boldsymbol{p}$ on J_m. In the tensor representation of the system and its solution we keep the temporal index as a single dimension (without quantization) connected to the first "virtual" spatial index, so that $\boldsymbol{P}_m$ is indexed by the tuples

$$\underbrace{j,}_{\text{time dim.}} \underbrace{j_{1,1},\ldots,j_{1,l_1}}_{\text{1st dimension}}, \underbrace{j_{2,1},\ldots,j_{2,l_2}}_{\text{2nd dimension}}, \ldots\ldots, \underbrace{j_{d,1},\ldots,j_{d,l_d}}_{d\text{th dimension}} \tag{21}$$

and

$$\underbrace{j,}_{\text{time dim.}} \underbrace{j_{1,1},\ldots,j_{d,1}}_{\text{1st level}}, \underbrace{j_{1,2},\ldots,j_{d,2}}_{\text{2nd level}}, \ldots\ldots, \underbrace{j_{1,l},\ldots,j_{d,l}}_{d\text{th level}} \tag{22}$$

in the QTT and QT3 formats respectively, cf. (13) and (15). The right-hand side of (20) is formed by attaching $\boldsymbol{\phi}_m$ to a QTT or QT3 decomposition of $\boldsymbol{p}_-^{m-1}$, therefore the first rank of the resulting decomposition is equal to 1 and the rest $\tilde{d}$ are the same as for $\boldsymbol{p}_-^{m-1}$. As for the matrix of (20), it can be trivially represented with the first rank equal to 2 and each remaining rank equal to 1 plus the corresponding rank of $\boldsymbol{A}$.

Theorem 6.4. *Assume that $\boldsymbol{A}$ is represented in the QTT or QT3 format in terms of $\tilde{d}$ cores with ranks $r_1, \ldots, r_{\tilde{d}-1}$. Then the matrix of system* (20) *can be represented in the corresponding format in terms of $\tilde{d}+1$ cores with ranks $2, r_1+1, \ldots, r_{\tilde{d}-1}+1$.*

The proof is given at the end of this supplement. □

6.3 Core matrices and the strong Kronecker product

By a *TT core* of rank $r_{k-1} \times r_k$ and mode size $m_k \times n_k$ we denote an array of real numbers, which has size $r_{k-1} \times m_k \times n_k \times r_k$. The first and the last indices of a core are called (respectively, *left* and *right*) *rank indices*, while the others are referred to as *mode indices*. Subarrays of a core, corresponding to particular values of rank indices, have size $m_k \times n_k$ and are called *TT blocks*. We may consider the core V_k as an $r_{k-1} \times r_k$-matrix with TT blocks as elements:

$$V_k = \begin{bmatrix} G_{11} & \cdots & G_{1r_k} \\ \vdots & \vdots & \vdots \\ G_{r_{k-1}1} & \cdots & G_{r_{k-1}r_k} \end{bmatrix} = \left[G_{\alpha_{k-1}\alpha_k}\right]_{\substack{\alpha_{k-1}=1,\ldots,r_{k-1} \\ \alpha_k=1,\ldots,r_k}}, \tag{23}$$

where $G_{\alpha_{k-1}\alpha_k}$, $\alpha_{k-1} = 1, \ldots, r_{k-1}$, $\alpha_k = 1, \ldots, r_k$ are TT blocks of V_k, i.e. $V_k(\alpha_{k-1}, i_k, j_k, \alpha_k) = \left(G_{\alpha_{k-1}\alpha_k}\right)_{i_k j_k}$ for all values of rank indices α_{k-1}, α_k and mode indices i_k, j_k. We refer to this matrix as *core matrix* of V_k.

In order to avoid confusion, we use parentheses for ordinary matrices, whose entries are numbers and which are multiplied as usual, and square brackets for cores (core matrices), whose entries are blocks and which are multiplied by means of the strong Kronecker product "$\bowtie$" defined below. Addition of cores is meant elementwise. Also, we may think of $G_{\alpha\beta}$ or of any submatrix of the core matrix in (23) as *subcores* of V_k. For example, given matrices or cores $U_{11}, U_{12}, U_{21}, U_{22}$ of equal mode size and compatible ranks, we may use them as subcores to compose the cores

$$\begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} U_{11} & \\ & U_{12} \end{bmatrix} = \operatorname{diag}\left[U_{11}, U_{22}\right].$$

We leave zero blocks blank, as in the last equation.

To ease notation, we omit in TT decompositions like (10), (11) the mode indices with the help of the *strong Kronecker product* [71]. To avoid the confusion with the Hadamard and tensor products, we denote this operation by "$\bowtie$", as in [52, Definition 2.1], where it was introduced as follows, specifically for connecting cores into "tensor trains".

Definition 6.5 (Strong Kronecker product $\bowtie$ of TT cores)**.** *Consider cores V_1 and V_2 of ranks $r_0 \times r_1$ and $r_1 \times r_2$ and of mode sizes $m_1 \times n_1$ and $m_2 \times n_2$ respectively, composed of blocks $G^{(1)}_{\alpha_0\alpha_1}$ and $G^{(2)}_{\alpha_1\alpha_2}$, $1 \le \alpha_k \le r_k$ for $0 \le k \le 2$. Then the* strong Kronecker product $V_1 \bowtie V_2$ *of V_1 and V_2 is defined as core of rank $r_0 \times r_2$ and mode size $m_1 m_2 \times n_1 n_2$, consisting of blocks*

$$G_{\alpha_0\alpha_2} = \sum_{\alpha_1=1}^{r_1} G^{(1)}_{\alpha_0\alpha_1} \otimes G^{(2)}_{\alpha_1\alpha_2}, \quad 1 \le \alpha_0 \le r_0, \quad 1 \le \alpha_2 \le r_2.$$

In other words, we define $V_1 \bowtie V_2$ as a usual matrix product of the corresponding core matrices, their entries (blocks) being multiplied by means of the Kronecker (tensor) product. For example,

$$\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \bowtie \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} G_{11} \otimes H_{11} + G_{12} \otimes H_{21} & G_{11} \otimes B_{12} + G_{12} \otimes H_{22} \\ G_{21} \otimes H_{11} + G_{22} \otimes H_{21} & G_{21} \otimes B_{12} + G_{22} \otimes H_{22} \end{bmatrix}.$$

Equation (11) can be written then as

$$\boldsymbol{A} = V_1 \bowtie V_2 \bowtie \ldots \bowtie V_{d-1} \bowtie V_d. \tag{24}$$

In the particular case when the second mode length is 1 in each core, the strong Kronecker product of them is a vector and the second mode indices can be omitted. For example, equation (11) reads

$$\boldsymbol{p} = U_1 \bowtie U_2 \bowtie \ldots \bowtie U_{d-1} \bowtie U_d. \tag{25}$$

6.4 Some operations in the TT format

In this section we present a few basic operations in the TT format. The results given for matrices are valid for vectors. Vice versa, the statements formulated for vectors hold for matrices too. Indeed, the latter can be vectorized by merging their mode indices, subjected to the operations in question, and the result can be turned back into a matrix.

Proposition 6.6 (Section 3.1 in[38]). *If a vector $\boldsymbol{p}$ is given in a CP decomposition*

$$\boldsymbol{p} = \sum_{\alpha=1}^{r} G_{1,\alpha} \otimes \ldots \otimes G_{d,\alpha},$$

it can be represented in the TT format as $\boldsymbol{p} = U_1 \bowtie \ldots \bowtie U_d$ with $U_k = \operatorname{diag}[G_{k,1}, \ldots, G_{k,r}]$ for $2 \le k \le d-1$,

$$U_1 = \begin{bmatrix} G_{1,1} & \ldots & G_{1,r} \end{bmatrix} \quad \text{and} \quad U_d = \begin{bmatrix} G_{d,1} \\ \vdots \\ G_{d,r} \end{bmatrix}.$$

In particular, the TT ranks are bounded by the CP rank.

Proposition 6.7 (Section 4.1 in[38]). *Assume that $\boldsymbol{p} = U_1 \bowtie \ldots \bowtie U_d$ and $\boldsymbol{q} = V_1 \bowtie \ldots \bowtie V_d$ are vectors of equal mode size, then a linear combination of them can be written as follows*

$$\alpha\boldsymbol{p} + \beta\boldsymbol{q} = \begin{bmatrix} U_1 & V_1 \end{bmatrix} \bowtie \operatorname{diag}[U_2, V_2] \bowtie \ldots \bowtie \operatorname{diag}[U_{d-1}, V_{d-1}] \bowtie \begin{bmatrix} \alpha U_d \\ \beta V_d \end{bmatrix}$$

for all $\alpha, \beta \in \mathbb{R}$.

Thus, the ranks of such a decomposition of $\alpha\boldsymbol{A} + \beta\boldsymbol{B}$ are sums of the corresponding ranks of the given decompositions of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Proposition 6.8. *If a vector $\boldsymbol{\omega}$ is given in the TT format through $\boldsymbol{\omega} = U_1 \bowtie \ldots \bowtie U_d$, then its diagonalization* $\operatorname{diag}\boldsymbol{\omega}$ *can be represented in the TT format as* $\operatorname{diag}\boldsymbol{\omega} = V_1 \bowtie \ldots \bowtie V_d$*, where the cores of the matrix are obtained by diagonalizing all the blocks in every core of the vector: $V_k(\alpha, i, j, \beta) = U_k(\alpha, j, \beta) \cdot \delta(i, j)$ for all α, i, j, β and for $1 \le k \le d$.*

Therefore, TT ranks are preserved under diagonalization.

Proposition 6.9 (Section 4.3 in[38]). *Consider matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ given in TT representations $\boldsymbol{A} = U_1 \bowtie \ldots \bowtie U_d$ and $\boldsymbol{B} = V_1 \bowtie \ldots \bowtie V_d$ of ranks $p_1, \ldots, p_{d-1}$ and $q_1, \ldots, q_{d-1}$ respectively. Let $p_0 = p_d = q_0 = q_d = 1$ and assume that for $1 \le k \le d$ the cores $U_k = \left[A_{\alpha_{k-1}\alpha_k}\right]_{\substack{\alpha_{k-1}=1,\ldots,p_{k-1} \\ \alpha_k=1,\ldots,p_k}}$ and $V_k = \left[B_{\beta_{k-1}\beta_k}\right]_{\substack{\beta_{k-1}=1,\ldots,q_{k-1} \\ \beta_k=1,\ldots,q_k}}$ are of such mode size that all matrix-matrix products $C_{\overline{\alpha_{k-1}\beta_{k-1}}\,\overline{\alpha_k\beta_k}} = A_{\alpha_{k-1}\alpha_k} \cdot B_{\beta_{k-1}\beta_k}$ are correctly defined. Then the matrix-matrix product $\boldsymbol{A} \cdot \boldsymbol{B}$ has a TT decomposition $\boldsymbol{A} \cdot \boldsymbol{B} = W_1 \bowtie \ldots \bowtie W_d$ with*

$$W_k = \left[C_{\overline{\alpha_{k-1}\beta_{k-1}}\,\overline{\alpha_k\beta_k}}\right]_{\substack{\alpha_{k-1}=1,\ldots,p_{k-1},\beta_{k-1}=1,\ldots,q_{k-1} \\ \alpha_k=1,\ldots,p_k,\beta_k=1,\ldots,q_k}}$$

and ranks $p_1q_1, \ldots, p_{d-1}q_{d-1}$.

The proof. The claim is obtained by writing the matrix-matrix product elementwise in terms of TT cores and in changing the summation order. □

For our considerations it is important that the corresponding TT ranks are multiplied under matrix-matrix multiplication.

Proposition 6.10. *Consider vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ given in TT decompositions $\boldsymbol{p} = U_1 \bowtie \ldots \bowtie U_d$ and $\boldsymbol{q} = V_1 \bowtie \ldots \bowtie V_d$. The tensor product $\boldsymbol{p} \otimes \boldsymbol{q}$ can be written as $\boldsymbol{p} \otimes \boldsymbol{q} = U_1 \bowtie \ldots \bowtie U_d \bowtie V_1 \bowtie \ldots \bowtie V_d$.*

In particular, the ranks of the first factor are followed by the ranks of the second factor with 1 in between. In what follows, we denote the operation of tensor transposition which was described in Section 2.2.4, by $\mathcal{T}$.

Lemma 6.11. *Consider vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ given in TT decompositions $\boldsymbol{p} = U_1 \bowtie \ldots \bowtie U_d$ and $\boldsymbol{q} = V_1 \bowtie \ldots \bowtie V_d$ of ranks $p_1, \ldots, p_{d-1}$ and $q_1, \ldots, q_{d-1}$ respectively. The transposed tensor product $\mathcal{T}(\boldsymbol{p} \otimes \boldsymbol{q})$ has a TT decomposition $\mathcal{T}(\boldsymbol{p} \otimes \boldsymbol{q}) = \overline{U}_1 \bowtie \overline{V}_1 \bowtie \overline{U}_2 \bowtie \overline{V}_2 \bowtie \ldots \bowtie \overline{U}_{d-1} \bowtie \overline{V}_{d-1} \bowtie \overline{U}_d \bowtie \overline{V}_d$ of ranks*

$$p_1, p_1 q_1,\ p_2 q_1, p_2 q_2,\ \ldots,\ p_{d-2} q_{d-2}, p_{d-1} q_{d-2},\ p_{d-1} q_{d-1},\ q_{d-1}$$

with $\overline{U}_1 = U_1$, $\overline{V}_d = V_d$ and the other cores defined as follows:

$$\begin{aligned} \overline{V}_1\big(\zeta_1, j_1, \overline{\eta_1 \beta_1}\big) &= V_1(j_1, \beta_1) \cdot \delta(\zeta_1, \eta_1)\,, \\ \overline{U}_d(\overline{\alpha_{d-1} \mu_{d-1}}, i_d, \nu_{d-1}) &= U_d(\alpha_{d-1}, i_d) \cdot \delta(\mu_{d-1}, \nu_{d-1}) \end{aligned}$$

and, for $2 \leq k \leq d-1$,

$$\begin{aligned} \overline{U}_k(\overline{\alpha_{k-1} \mu_{k-1}}, i_k, \overline{\alpha_k \nu_{k-1}}) &= U_k(\alpha_{k-1}, i_k, \alpha_k) \cdot \delta(\mu_{k-1}, \nu_{k-1})\,, \\ \overline{V}_k\big(\overline{\zeta_k \beta_{k-1}}, j_k, \overline{\eta_k \beta_k}\big) &= V_k(\alpha_{k-1}, j_k, \alpha_k) \cdot \delta(\zeta_k, \eta_k) \end{aligned}$$

for all mode indices i_k, j_k, where $1 \leq k \leq d$, and for $1 \leq \alpha_k, \zeta_k, \eta_k \leq p_k$ and $1 \leq \beta_k, \mu_k, \nu_k \leq q_k$, where $1 \leq k \leq d-1$.

Proof. By changing the order of summation and multiplication, for all values of mode indices $i_1, \ldots, i_d$, $j_1, \ldots, j_d$ we obtain

$$\begin{aligned} &(\boldsymbol{p} \otimes \boldsymbol{q})_{\,i_1,\ldots,i_d,j_1,\ldots,j_d} \\ &= \sum_{\alpha_1=1}^{p_1} \ldots \sum_{\alpha_{d-1}=1}^{p_{d-1}} U_1(i_1, \alpha_1) \cdot U_2(\alpha_1, i_2, \alpha_2) \cdot \ldots \\ &\qquad \cdot U_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) \cdot U_d(\alpha_{d-1}, i_d) \\ &\quad \cdot \sum_{\beta_1=1}^{q_1} \ldots \sum_{\beta_{d-1}=1}^{q_{d-1}} V_1(j_1, \beta_1) \cdot V_2(\beta_1, j_2, \beta_2) \cdot \ldots \\ &\qquad \cdot V_{d-1}(\beta_{d-2}, j_{d-1}, \beta_{d-1}) \cdot V_d(\beta_{d-1}, j_d) \\ &= \sum_{\alpha_1=1}^{p_1} \sum_{\beta_1=1}^{q_1} \ldots \sum_{\alpha_{d-1}=1}^{p_{d-1}} \sum_{\beta_{d-1}=1}^{q_{d-1}} U_1(i_1, \alpha_1) \cdot V_1(j_1, \beta_1) \\ &\qquad \cdot U_2(\alpha_1, i_2, \alpha_2) \cdot V_2(\beta_1, j_2, \beta_2) \cdot \ldots \\ &\quad \cdot U_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) \cdot V_{d-1}(\beta_{d-2}, j_{d-1}, \beta_{d-1}) \cdot \\ &\qquad \cdot U_d(\alpha_{d-1}, i_d) \cdot V_d(\beta_{d-1}, j_d) \\ &= (\mathcal{T}(\boldsymbol{p} \otimes \boldsymbol{q}))_{\,i_1,j_1,\ldots,i_d,j_d}. \end{aligned}$$

□

Lemma 6.12. *Let U be a core of rank $p_0 \times p_d$ with a d-dimensional mode index. For $r \in \mathbb{N}$ consider the core $\overline{U}$ defined by setting*

$$\overline{U}\big(\overline{\alpha_0\gamma_0}, \overline{i_1, \ldots, i_d}, \overline{\alpha_d\gamma_d}\big) = U\big(\alpha_0, \overline{i_1, \ldots, i_d}, \alpha_d\big) \cdot \delta(\gamma_0, \gamma_d)$$

for all values of mode indices $i_1, \ldots, i_d$, for $1 \le \alpha_0 \le p_0$, $1 \le \alpha_d \le p_d$ and for $1 \le \gamma_0, \gamma_d \le r$.

Assume that U is given in a decomposition $U = U_1 \bowtie U_2 \bowtie \ldots \bowtie U_{d-1} \bowtie U_d$, where U_k is of rank $p_{k-1} \times p_k$. Then $\overline{U}$ can be represented as $\overline{U} = \overline{U}_1 \bowtie \overline{U}_2 \bowtie \ldots \bowtie \overline{U}_{d-1} \bowtie \overline{U}_d$, where for $1 \le k \le d$ the core $\overline{U}_k$ of rank $p_{k-1}r \times p_k r$ is defined as follows:

$$\overline{U}_k(\overline{\alpha_{k-1}\gamma_{k-1}}, i_k, \overline{\alpha_k\gamma_k}) = U_k(\alpha_{k-1}, i_k, \alpha_k) \cdot \delta(\gamma_{k-1}, \gamma_k)$$

for all values of mode index i_k, for $1 \le \alpha_{k-1} \le p_{k-1}$, $1 \le \alpha_k \le p_k$ and for $1 \le \gamma_{k-1}, \gamma_k \le r$.

Proof. For all rank and mode indices we have

$$\begin{aligned}
&\overline{U}\big(\overline{\alpha_0\gamma_0}, \overline{i_1, \ldots, i_d}, \overline{\alpha_d\gamma_d}\big) \\
&= \sum_{\alpha_1=1}^{p_1} \ldots \sum_{\alpha_{d-1}=1}^{p_{d-1}} \delta(\gamma_0, \gamma_d) \prod_{k=1}^{d} U_k(\alpha_{k-1}, i_k, \alpha_k) \\
&= \sum_{\gamma_1=1}^{r} \sum_{\alpha_1=1}^{p_1} \ldots \sum_{\gamma_{d-1}=1}^{r} \sum_{\alpha_{d-1}=1}^{p_{d-1}} \prod_{k=1}^{d} \delta(\gamma_{k-1}, \gamma_k)\, U_k(\alpha_{k-1}, i_k, \alpha_k)\,, \qquad (26)
\end{aligned}$$

i.e. $\overline{U} = \overline{U}_1 \bowtie \overline{U}_2 \bowtie \ldots \bowtie \overline{U}_{d-1} \bowtie \overline{U}_d$. □

Corollary 6.13. *Assume that vectors $\boldsymbol{p}_k$, $1 \le k \le d$, are given in QTT decompositions with l quantization levels and of ranks $r_1^{(k)}, \ldots, r_{l-1}^{(k)}$, $1 \le k \le d$, respectively. Then the tensor product $\boldsymbol{p}_1 \otimes \ldots \otimes \boldsymbol{p}_d$ can be represented in the transposed QTT format with ranks*

$$\begin{gathered}
r_1^{(1)},\ r_1^{(1)}r_1^{(2)},\ \ldots,\ r_1^{(1)} \cdot r_1^{(2)} \cdot \ldots \cdot r_1^{(d-1)} \cdot r_1^{(d)}, \\
\color{blue}{r_1^{(1)}r_1^{(2)} \cdot \ldots \cdot r_1^{(d-1)} \cdot r_1^{(d)}}, \\
r_2^{(1)} \cdot r_1^{(2)} \cdot \ldots \cdot r_1^{(d)} \cdot r_1^{(d)},\ \ldots,\ r_2^{(1)} \cdot r_2^{(2)} \cdot \ldots \cdot r_2^{(d)} \cdot r_1^{(d)}, \\
\color{blue}{r_2^{(1)}r_2^{(2)} \cdot \ldots \cdot r_2^{(d-1)} \cdot r_2^{(d)}}, \\
\ldots\ldots \\
\color{blue}{r_{l-2}^{(1)}r_{l-2}^{(2)} \cdot \ldots \cdot r_{l-2}^{(d-1)} \cdot r_{l-2}^{(d)}}, \\
r_{l-1}^{(1)} \cdot r_{l-2}^{(2)} \cdot \ldots \cdot r_{l-2}^{(d)} \cdot r_{l-2}^{(d)},\ \ldots,\ r_{l-1}^{(1)} \cdot r_{l-1}^{(2)} \cdot \ldots \cdot r_{l-1}^{(d-1)} \cdot r_{l-2}^{(d)}, \\
\color{blue}{r_{l-1}^{(1)}r_{l-1}^{(2)} \cdot \ldots \cdot r_{l-1}^{(d-1)} \cdot r_{l-1}^{(d)}}, \\
r_{l-1}^{(2)} \cdot \ldots \cdot r_{l-1}^{(d)},\ \ldots,\ r_{l-1}^{(d-1)} \cdot r_{l-1}^{(d)},\ r_{l-1}^{(d)},
\end{gathered}$$

where we highlight in blue the QT3 ranks separating adjacent levels. As a result, all QT3 ranks of the tensor product are bounded from above by

$$\prod_{k=1}^{d} \max_{1 \le m \le l-1} r_m^{(k)}.$$

The proof follows from Lemmas 6.11 and 6.12. □

When the numbers of levels of quantization vary, i.e. $l_1 = \ldots = l_d = l$ does not hold for any l, Corollary 6.13 still remains true. Indeed, one can increase the number of cores in each decomposition up to $\max_{1 \le k \le d} l_k$ by introducing void cores with void mode indices and of rank 1×1 (so that these cores are just constant factors), apply the results presented above and remove the void cores by contracting them with the others.

6.5 Proofs of the theorems

Proof of Theorem 2.4. Assume $1 \le s \le R$. The corresponding shift matrix is a Kronecker product:

$$\boldsymbol{S}_{\underline{\eta}^s} = \boldsymbol{S}^{(l_1)}_{\eta^s_1} \otimes \ldots \otimes \boldsymbol{S}^{(l_d)}_{\eta^s_d},$$

where $\boldsymbol{S}^{(l_k)}_{\eta_k}$ is the $2^{l_k} \times 2^{l_k}$-matrix of the zero-filling η^s_k-position shift, $1 \le k \le d$. By [53, Lemma 4.2], each of these one-dimensional factors can be represented explicitly in the QTT format with ranks bounded by 2 for any η^s_k. However, if $\eta^s_k = 0$, then $\boldsymbol{S}^{(l_k)}_{\eta_k} = \mathbb{I}$ is of QTT ranks $1, \ldots, 1$. Therefore, according to Proposition 6.10, the QTT ranks of $\boldsymbol{S}_{\underline{\eta}^s}$ are bounded by $\rho^s_1, \ldots, \rho^s_1$, $1, \rho^s_2, \ldots, \rho^s_2$, $1, \ldots, \ldots, 1, \rho^s_d, \ldots, \rho^s_d$, where

$$\rho^s_k = \begin{cases} 2, & \eta^s_k \neq 0, \\ 1, & \eta^s_k = 0 \end{cases}$$

for $1 \le k \le d$. As the identity matrix is of QTT ranks $1, \ldots, 1$, by Proposition 6.7, the QTT ranks of $\boldsymbol{S}_{\underline{\eta}^s} - \mathbb{I}$ are bounded by $\rho^s_1+1, \ldots, \rho^s_1+1$, $2, \rho^s_2+1, \ldots, \rho^s_2+1$, $2, \ldots, \ldots, 2, \rho^s_d+1, \ldots, \rho^s_d+1$.

Analogously we obtain that the QTT ranks of $\boldsymbol{\omega}^s$ are bounded by $r^s_1, \ldots, r^s_1$, $1, r^s_2, \ldots, r^s_2$, $1, \ldots, \ldots, 1$, $r^s_d, \ldots, r^s_d$, where for $1 \le k \le d$ we have $r^s_k = 1$ if $\eta^s_k = 0$. Due to Proposition 6.8, the same bounds hold true for the QTT ranks of the matrix $\operatorname{diag} \boldsymbol{\omega}^s$.

Finally, we use Proposition 6.9 to conclude that the sth term $\left(\boldsymbol{S}_{\underline{\eta}^s} - \mathbb{I}\right) \circ \boldsymbol{M}_{\omega^s}$ of the CME operator is represented in the QTT format with ranks bounded by $\tilde{q}^s_1, \ldots, \tilde{q}^s_1$, $1, \tilde{q}^s_2, \ldots, \tilde{q}^s_2$, $1, \ldots, \ldots, 1, \tilde{q}^s_d, \ldots, \tilde{q}^s_d$, where

$$\tilde{q}^s_k = \begin{cases} 3 \cdot r^s_k, & \eta^s_k \neq 0, \\ 2 \cdot 1, & \eta^s_k = 0 \end{cases}$$

for $1 \le k \le d$. By summing these rank bounds, we obtain the rank bounds claimed for $\boldsymbol{A}$ with $q_k = \sum_{s=1}^{R} \tilde{q}^s_k$. □

Proof of Theorem 2.5. Analogous to that of Theorem 2.4. For the QT3 format, we use Corollary 6.13 instead of Proposition 6.10 to construct tensor products and to establish the corresponding rank bounds. □

Proof of Theorem 6.4. Let us set $U_0 = [\boldsymbol{C}_m]$, $U_m = [\mathbb{I}_{\rho_m+1}]$ for $1 \le m \le \sum_{k=1} l_k$, $V_0 = [-\boldsymbol{G}_m]$ and assume $\boldsymbol{A} = U_1 \bowtie \ldots \bowtie U_{\sum_{k=1} l_k}$. Then the proof follows from Proposition 6.7.

□

Recent Research Reports

Nr.	Authors/Title

2012-37 C. Schillings and C. Schwab
Sparse, adaptive Smolyak algorithms for Bayesian inverse problems

2012-38 R. Hiptmair and A. Moiola and I. Perugia and C. Schwab
Approximation by harmonic polynomials in star-shaped domains and exponential convergence of Trefftz hp-DGFEM

2012-39 A. Buffa and G. Sangalli and Ch. Schwab
Exponential convergence of the hp version of isogeometric analysis in 1D

2012-40 D. Schoetzau and C. Schwab and T. Wihler and M. Wirz
Exponential convergence of hp-DGFEM for elliptic problems in polyhedral domains

2012-41 M. Hansen
n-term approximation rates and Besov regularity for elliptic PDEs on polyhedral domains

2012-42 C. Gittelson and R. Hiptmair
Dispersion Analysis of Plane Wave Discontinuous Galerkin Methods

2012-43 J. Waldvogel
Jost Bürgi and the discovery of the logarithms

2013-01 M. Eigel and C. Gittelson and C. Schwab and E. Zander
Adaptive stochastic Galerkin FEM

2013-02 R. Hiptmair and A. Paganini and M. Lopez-Fernandez
Fast Convolution Quadrature Based Impedance Boundary Conditions

2013-03 X. Claeys and R. Hiptmair
Integral Equations on Multi-Screens