Solving chemical master equations by an adaptive wavelet method

Tobias Jahnke* and Steffen Galan†

*Universität Karlsruhe (TH), Fakultät für Mathematik, Institut für Angewandte und Numerische Mathematik, Englerstr. 2, D-76128 Karlsruhe, Germany, jahnke@math.uni-karlsruhe.de
†galan@math.fu-berlin.de

Abstract. Solving chemical master equations is notoriously difficult due to the tremendous number of degrees of freedom. We present a new numerical method which efficiently reduces the size of the problem in an adaptive way. The method is based on a sparse wavelet representation and an algorithm which, in each time step, detects the essential degrees of freedom required to approximate the solution up to the desired accuracy.

Keywords: Stochastic reaction kinetics, chemical master equation, Haar wavelets, adaptive Galerkin method, Rothe method
PACS: 02.60.Lj, 02.70.Dh, 82.39.Rt, 05.10.Gg, 87.10.Ed, 87.18.Vf

STOCHASTIC REACTION KINETICS

The traditional reaction-rate approach often fails to model intracellular reaction systems correctly. In gene regulatory networks, signalling pathways, or viral replication, some of the substances are present in such a low number of molecules that it does not make sense to describe their occurrence in terms of continuously varying concentrations. Moreover, the behaviour of the entire system can be crucially changed by small stochastic fluctuations which are neglected by the deterministic reaction-rate approach. In order to account for the discrete and random nature of intracellular processes the problem must be formulated in terms of stochastic reaction kinetics. The system is then represented by a time-dependent probability distribution

$$p(t,\,\cdot\,) \;\in\; l^1 \;=\; \Big\{ f : \Omega \to \mathbb{R} \text{ with } \sum_{x\in\Omega} |f(x)| < \infty \Big\}$$

on the d-dimensional state space $\Omega \subset \mathbb{N}^d$. The entries of a state $x = (x_1, \ldots, x_d) \in \Omega$ are the number of copies of each species such that $p(t,x)$ is the probability that at time t there are exactly x_i copies of the i-th species ($i = 1, \ldots, d$). It is well-known that p is a solution of the *chemical master equation* (CME)

$$\partial_t p(t,\cdot) \quad = \quad Ap(t,\,\cdot\,) \tag{1}$$

where A denotes the operator

$$\big(Ap(t,\,\cdot\,)\big)(x) \quad = \quad \sum_{m=1}^{M} \big(\alpha_m(x-\nu_m)p(t,x-\nu_m) - \alpha_m(x)p(t,x)\big) \tag{2}$$

on a suitable dense domain $D(A) \subset l^1$ (cf. [1, 2, 3]). Here, ν_m is the stoichiometric vector of the m-th reaction and $\alpha_m(x) \geq 0$ denotes the corresponding propensity function. For simplicity of notation, we let $p(t,x) = 0$ and $\alpha_m(x) = 0$ for all $x \notin \Omega$. Of course, Eq. (1) has to be complemented by an initial condition $p(0,x) = p_0(x)$ for some $p_0 \in l^1$.

Although Ω is typically infinite, numerical computations can only be performed on a truncated state space

$$\Omega_\xi = \{x \in \mathbb{N}^d \,:\, x_1 < \xi_1,\, \ldots \,,\, x_d < \xi_d\},$$

where $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{N}^d$ is some suitably chosen truncation vector. The error caused by truncating the state space has been investigated in [4]. Here, we simply assume that Ω_ξ is chosen large enough such that the truncation error can be neglected because $p(t,\,\cdot\,)$ vanishes near the artificial boundary. Of course, the truncation vector ξ could be chosen adaptively, but for simplicity ξ will be kept fixed in what follows.

The truncated (finite) state space contains $N = \xi_1 \cdot \ldots \cdot \xi_d$ states, and each state $x \in \mathbb{R}^{\xi_1 \times \cdots \times \xi_d}$ corresponds to one degree of freedom in the solution $p(t, \cdot)$. In typical applications the dimension d and the upper limits ξ_i are so large that any attempt to solve the CME with traditional techniques is absolutely hopeless. Therefore, many applications in computational biology are based on Monte Carlo simulations using the *stochastic simulation algorithm* (cf. [1, 5]) or modifications thereof, but this algorithm requires an update whenever one of the reaction channels fires, and thus sampling a full time-dependent probability distribution can be computationally costly. Only during the last few years a number of approaches to solve the CME directly have been proposed, cf. [4, 6, 7, 8, 9, 10, 11, 12, 13, 14].

In this short note we devise an adaptive Galerkin method based on a sparse, adaptive wavelet representation of the solution. Since the time-dependent probability distribution $p(t,\cdot)$ is locally "smooth" (in a discrete sense), the wavelet compression allows to decrease the size of the problem very efficiently such that the remaining, essential degrees of freedom can be propagated by the Rothe method. The wavelet approach is particularly suited for the CME because the geometry of the state space is simple, and the boundary conditions can easily be incorporated into the definition of the CME operator. The crucial problem, however, is the question how the essential basis elements can be determined in each time step. We show how this can be done in an efficient and reliable way. In this note, only Haar wavelets are considered; an extension to more sophisticated wavelets is currently investigated.

ADAPTIVE GALERKIN METHOD WITH SPARSE WAVELET BASIS

Let $\mathscr{H}(\Omega_\xi)$ be the space of all discrete functions $f : \Omega_\xi \longrightarrow \mathbb{R}$ with inner product

$$\langle f\ ,\ g\rangle = \sum_{x\in\Omega_\xi} f(x)g(x), \qquad f,g \in \mathscr{H}(\Omega_\xi). \tag{3}$$

Suppose that $N = \xi_1 \cdot \ldots \cdot \xi_d$ is the total number of states, and let $\{\psi_1, \ldots, \psi_N\}$ denote the d-dimensional Haar wavelet basis (see, e.g., [15]). Moreover, let

$$\Psi : \mathbb{R}^N \longrightarrow \mathscr{H}, \qquad \Psi : (a_1,\ldots,a_N) \longmapsto \sum_{j=1}^{N} a_j\psi_j \tag{4}$$

be the linear operator which maps a coefficient vector to the corresponding element in $\mathscr{H}$. Since the wavelet basis is orthonormal, the operator Ψ is orthogonal and therefore an isometry. The adjoint is given by

$$\Psi^* : \mathscr{H} \longrightarrow \mathbb{R}^N, \qquad \Psi^* : p \longmapsto (\langle\psi_1\ ,\ p\rangle, \ldots, \langle\psi_N\ ,\ p\rangle).$$

Both operators Ψ and Ψ^* can efficiently be applied by fast wavelet transforms.

The motivation for using the wavelet representation is the observation that many coefficients of a sufficiently smooth $p \in \mathscr{H}$ almost vanish and thus can be neglected if a small approximation error is permitted (cf. [15]). Suppose that the basis is ordered in such a way that the initial distribution p_0 can be approximately represented by the first $\eta \le N$ basis elements, i.e.

$$p_0 \approx y_0 = \sum_{j=1}^{\eta} a_j\psi_j.$$

Then, the problem can be projected to the low-dimensional subspace $span\{\psi_1,\ldots,\psi_\eta\}$ by the Galerkin condition

$$\langle \partial_t y - Ay\ ,\ q\rangle = 0 \qquad \text{for all} \quad q \in span\{\psi_1,\ldots,\psi_\eta\}.$$

The evolution of the coefficients $a(t) = (a_1(t), \ldots, a_\eta)(t)$ corresponding to $y(t)$ is given by the differential equation

$$\frac{da}{dt} = A_\eta a \qquad \text{with} \quad A_\eta = \Big(\big\langle\psi_i\ ,\ A\psi_j\big\rangle\Big)_{i,j}. \tag{5}$$

A traditional, non-adaptive Galerkin method (also known as method of lines) would now proceed by first solving (5) numerically and then computing the linear combination $y(t) = \sum_{j=1}^{\eta} a_j(t)\psi_j$. If $\eta \ll N$, then the Galerkin approximation y can be computed with considerably lower efforts because the number of degrees of freedom is reduced from N to

η. However, the approximation error $\|y(t,\cdot) - p(t,\cdot)\|$ can become prohibitively large as t increases, because the non-adaptive Galerkin method is confined to the small subspace $span\{\psi_1,\ldots,\psi_\eta\}$ whereas the exact solution propagates in the entire space $\mathscr{H}$.

This motivates the key idea of adaptive Galerkin methods: to change the subset of basis elements as time evolves such that the solution can always be sufficiently well approximated. Let $h > 0$ be the step size and suppose that

$$p(t_n) \approx y_n = \sum_{i=1}^{\eta} a_i \psi_{j_i} \tag{6}$$

where $t_n = t_0 + nh$ and $\{j_1, \ldots, j_\eta\}$ is a subset of the index set $\{1, \ldots, N\}$. Our task is to find a new selection $\{k_1, \ldots, k_\mu\} \subset \{1, \ldots, N\}$ and new coefficients b_i such that

$$p(t_{n+1}) \approx y_{n+1} = \sum_{i=1}^{\mu} b_i \psi_{k_i}.$$

Of course, the number of coefficients μ is supposed to be as small as possible. In other words, our goal is in each time-step to identify the essential degrees of freedom and to propagate the solution in the corresponding basis.

To this end we apply the Rothe method (cf., e.g., [14] and references therein): In contrast to the method of lines which first projects the problem into a fixed approximation subspace and then performs a time discretisation, the Rothe method first discretises the problem in time and then adapts the spatial approximation in each time step. Let $p_{n+1} \approx p(t_{n+1})$ denote the approximation obtained by the trapezoidal rule applied to the (truncated) CME (1). For given p_n, the new approximation p_{n+1} is the solution of the linear system

$$\Big(I - \frac{h}{2}A\Big) p_{n+1} \quad = \quad \Big(I + \frac{h}{2}A\Big) p_n. \tag{7}$$

This system contains all N degrees of freedom and is therefore too large to be solved. However, it does not make sense to solve (7) *exactly* because the time-discretisation has already caused a small approximation error. Therefore, it is sufficient to approximate $p_{n+1} \approx y_{n+1}$ up to a given tolerance `tol` and to determine the k_i and b_i such that

$$\Big\| \Big(I - \frac{h}{2}A\Big) y_{n+1} - \Big(I + \frac{h}{2}A\Big) y_n \Big\| \le \texttt{tol} \qquad \text{holds for} \qquad y_{n+1} = \sum_{i=1}^{\mu} b_i \psi_{k_i}. \tag{8}$$

If the new selection of basis elements is known, the coefficient vector $b = (b_i)_i$ can easily be computed by projecting the system (7) into the corresponding (low-dimensional) space $span\{\psi_{k_i} \mid i = 1,\ldots,\mu\}$. This yields

$$Mb \quad = \quad f \qquad \text{with} \qquad \begin{array}{rclcrcl} M & = & (m_{il})_{i,l}\,, & & m_{il} & = & \big\langle \psi_{k_i}\,,\, (I - \frac{h}{2}A)\psi_{k_l} \big\rangle, \\ f & = & (f_i)_i\,, & & f_i & = & \big\langle \psi_{k_i}\,,\, (I + \frac{h}{2}A) y_n \big\rangle. \end{array} \tag{9}$$

SELECTING NEW BASIS ELEMENTS

The main problem, however, remains: *how can we identify the essential basis elements* in each time-step, i.e., how do we select $\{k_1, \ldots, k_\mu\}$? Clearly, it would be highly inefficient to enlarge the basis randomly until (8) is met. Let $y_n = \sum_{i=1}^{\eta} a_i \psi_{j_i}$ be the approximation at time $t = t_n$, and let Φ be the orthogonal projection onto the subspace $span\{\psi_{j_1},\ldots,\psi_{j_\eta}\}$. Then, the Galerkin approximation y on the next time interval $[t_n, t_{n+1}]$ is the solution of $\partial_t y(t,\cdot) = \Phi A y(t,\,\cdot\,)$ with initial value $y(t_n,\,\cdot\,) = y_n$. Comparing with the CME (1) on the full space shows that the error $d(t,\,\cdot\,) = p(t,\cdot) - y(t,\cdot)$ evolves according to

$$\partial_t d(t,\cdot) \quad = \quad A d(t,\,\cdot\,) + (I - \Phi) A y(t,\,\cdot\,), \qquad d(t_n,\,\cdot\,) = 0.$$

In order to improve the Galerkin approximation the basis has to be enlarged by those elements ψ_l where

$$\big|\, \langle \psi_l\,,\, \partial_t d(t_n,\,\cdot\,) \rangle \,\big| = \big|\, \langle \psi_l\,,\, (I - \Phi) A y_n \rangle \,\big|$$

is particularly large. This leads to the following iterative strategy:

Input: index subset $\{j_1, \ldots, j_\eta\}$ and coefficients $a_1, \ldots, a_\eta$ of the current approximation $y_n = \sum_{i=1}^{\eta} a_i \psi_{j_i}$.

Output: index subset $\{k_1, \ldots, k_\mu\}$ and coefficients $b_1, \ldots, b_\mu$ of the new approximation $y_{n+1} = \sum_{i=1}^{\mu} b_i \psi_{k_i}$.

1. Set $\mu = \eta$ and $k_i = j_i$ for all $i = 1, \ldots, \eta$.
2. Solve the linear system (9) and obtain the coefficients $b_1, \ldots, b_\mu$.
3. Compute the new approximation $y_{n+1} = \sum_{i=1}^{\mu} b_i \psi_{k_i}$.
4. Compute the residual $r = \left(I - \frac{h}{2}A\right) y_{n+1} - \left(I + \frac{h}{2}A\right) y_n$
5. If $\|r\| > \texttt{tol}_1$:
   (a) Compute $\gamma = \Psi^*(I - \Phi)Ay_n$ by a fast wavelet transform.
   (b) Choose a number $\Delta\mu \in \mathbb{N} \setminus \{0\}$ and find the indices $k_{\mu+1}, \ldots, k_{\mu+\Delta\mu}$ of the $\Delta\mu$ largest entries of $|\gamma|$.
   (c) Enlarge the basis: add $\psi_{k_{\mu+1}}, \ldots, \psi_{k_{\mu+\Delta\mu}}$ to the current set of basis elements and put $\mu \mapsto \mu + \Delta\mu$
   (d) Go to step 2.
6. Remove all basis elements corresponding to small coefficients as long as the truncation error remains smaller than $\texttt{tol}_2$. Since the wavelet transform is an isometry, the truncation error is simply the 2-norm of the removed coefficients.

Remark 1. Rather than computing p_n, p_{n+1}, and r on the entire state space, it may be more efficient to compute these terms only on the essential support, i.e. on the subset of the state space where $p_n(x)$ is larger than some threshold.

Remark 2. The definition of γ in step 5a can be replaced by $\gamma = \Psi^*(I - \Phi)A(y_n + y_{n+1})/2$; this is often a better indicator for detecting essential basis elements.

Remark 3. The linear system (9) is solved by a LU-decomposition. If new basis elements are added, new columns and rows are appended to the matrix M in (9), but the LU decomposition can be updated and does not have to be computed from scratch.

This algorithm has been applied, e.g., to a stochastic SEIS model describing the infection of a population by a virus. The promising numerical results showed the potential of the approach to identify the essential degrees of freedom and to decrease the computational costs significantly, and we believe that the efficiency of our method can be considerably improved by further refinement of the approach.

REFERENCES

1. D. T. Gillespie, *J. Comput. Phys.* **22**, 403–434 (1976).
2. D. T. Gillespie, *Physica A* **188**, 404–425 (1992).
3. C. W. Gardiner, *Handbook of Stochastic Methods*, Springer, Berlin, 1985, 2nd enlarged edition edn.
4. B. Munsky, and M. Khammash, *J. Chemical Physics* (2006).
5. D. T. Gillespie, *J. Phys. Chem.* **81**, 2340–2361 (1977).
6. K. Burrage, M. Hegland, S. MacNamara, and R. B. Sidje, "A Krylov-based Finite State Projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems," in *Markov Anniversary Meeting: An international conference to celebrate the 150th anniversary of the birth of A.A. Markov*, edited by A.N.Langville, and W.J.Stewart, Boson Books, 2006, pp. 21 – 38.
7. S. MacNamara, K. Burrage, and R. B. Sidje, Multiscale modeling of chemical kinetics via the master equation (2007), unpublished.
8. M. Hegland, C. Burden, L. Santoso, S. MacNamara, and H. Booth, *Journal of Computational and Applied Mathematics* **205**, 708–724 (2007).
9. S. Engblom, A discrete spectral method for the chemical master equation, Tech. Rep. Technical Report 2006-036, Uppsala University (2006).
10. A. Hellander, and P. Lötstedt, *J. Comput. Phys.* **227**, 100–122 (2007).
11. S. Peles, B. Munsky, and M. Khammash, *Journal of Chemical Physics* **125**, 204104 (2006).
12. T. Jahnke, and W. Huisinga, *Journal of Mathematical Biology* **54**, 1–26 (2007).
13. T. Jahnke, and W. Huisinga, *Bulletin of Mathematical Biology* (2008).
14. P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow, Adaptive discrete Galerkin methods applied to the chemical master equation (2007), accepted for publication in SISC.
15. W. Dahmen, "Multiscale and wavelet methods for operator equations," in *Multiscale problems and methods in numerical simulations. Lectures given at the C.I.M.E. summer school, Martina Franca, Italy, September 9–15, 2001.*, edited by C. Canuto, Lect. Notes Math. 1825, Springer., Berlin, 2003, pp. 31–96.