

MATHEMATICAL MODELLING AND ANALYSIS
Volume 20 Number 3, May 2015, 382–395
<http://dx.doi.org/10.3846/13926292.2015.1048760>
© Vilnius Gediminas Technical University, 2015

Publisher: Taylor&Francis and VGTU
<http://www.tandfonline.com/TMMA>
ISSN: 1392-6292
eISSN: 1648-3510

Approximate Exponential Algorithms to Solve the Chemical Master Equation

Azam Mooasvi and Adrian Sandu

*Computational Science Laboratory, Department of Computer Science,
Virginia Polytechnic Institute and State University*

Blacksburg, VA 24060, USA

E-mail(*corresp.*): azmosavi@vt.edu

E-mail: sandu@cs.vt.edu

Received October 17, 2014; revised April 13, 2015; published online May 15, 2015

Abstract. This paper discusses new simulation algorithms for stochastic chemical kinetics that exploit the linearity of the chemical master equation and its matrix exponential exact solution. These algorithms make use of various approximations of the matrix exponential to evolve probability densities in time. A sampling of the approximate solutions of the chemical master equation is used to derive accelerated stochastic simulation algorithms. Numerical experiments compare the new methods with the established stochastic simulation algorithm and the tau-leaping method.

Keywords: stochastic chemical kinetics, chemical master equation, exact solution, stochastic simulation algorithm, tau-leap.

AMS Subject Classification: 92Bxx, 97Mxx.

1 Introduction

In many biological systems the small number of participating molecules make the chemical reactions inherently stochastic. The system state is described by probability densities of the numbers of molecules of different species. The evolution of probabilities in time is described by the chemical master equation (CME) [8]. Gillespie proposed the Stochastic Simulation Algorithm (SSA), a Monte Carlo approach that samples from CME [8]. SSA became the standard method for solving well-stirred chemically reacting systems. However, SSA simulates one reaction and is inefficient for most realistic problems. This motivated the quest for approximate sampling techniques to enhance the efficiency.

The first approximate acceleration technique is the tau-leaping method [9] which is able to simulate multiple chemical reactions appearing in a pre-selected time step of length τ . The tau-leap method is accurate if τ is small enough to satisfy the leap condition, meaning that propensity functions remain nearly constant in a time step. The number of firing reactions in a time step is approximated by a Poisson random variable [10]. Explicit tau-leaping method is

numerically unstable for stiff systems [5]. Stiffness systems have well-separated “fast” and “slow” time scales present, and the “fast modes” are stable. The implicit tau-leap method [12] overcomes the stability issue but it has a damping effect on the computed variances. More accurate variations of the implicit tau-leap method have been proposed to alleviate the damping [1, 3, 4, 7, 9, 13]. Simulation efficiency has been increased via parallelization [2].

Direct solutions of the CME are computationally important specially in order to estimate moments of the distributions of the chemical species [11]. Various approaches to solve the CME are discussed in [6].

Sandu re-drives the SSA algorithm from CME in [14]. The analysis reveals the hidden approximations made by the SSA and the tau-leap methods. The approach explains the explicit tau-leap method as an exact sampling procedure from an approximate solution of the CME. The numerical solver is obtained by first discretizing the CME in time, advancing the probability density by one time step, and then sampling the new probability density from the approximation.

This paper extends the study [14] and proposes new approximations to the CME solution based on various approximations of matrix exponentials. Different approximation techniques of the matrix exponential lead to various algorithms that have their own strengths and weaknesses with regard to accuracy, stability and computational complexity. Numerical experiments are performed with two different chemical systems to assess the accuracy and stability of each of the algorithms.

The paper is organized as follows. Section 2 reviews the stochastic simulation of chemical kinetics. Section 3 developed the new approximation methods. Numerical experiments to illustrate the proposed schemes are carried out in Section 4. Conclusions are drawn in Section 5.

2 Simulation of Stochastic Chemical Kinetics

Consider a chemical system in a constant volume container. The system is well-stirred and in thermal equilibrium at some constant temperature. There are N different chemical species S^1, \dots, S^N . Let $X^i(t)$ denote the number of molecules of species S_i at time t . The state vector $x(t) = [X^1(t), \dots, X^N(t)]$ defines the numbers of molecules of each species present at time t . The chemical network consists of M reaction channels R_1, \dots, R_M . Each individual reaction destroys a number of molecules of reactant species, and produces a number of molecules of the products. Let ν_j^i be the change in the number of S^i molecules caused by a single reaction R_j . The state change vector $\nu_j = [\nu_j^1, \dots, \nu_j^N]$ describes the change in the entire state following R_j .

A propensity function $a_j(x)$ is associated with each reaction channel R_j . The probability that one R_j reaction will occur in the next infinitesimal time interval $[t, t+dt)$ is $a_j(x(t)) \cdot dt$. The purpose of a stochastic chemical simulation is to trace the time evolution of the system state $x(t)$ given that at the initial time \bar{t} the system is in the initial state $x(\bar{t})$.

2.1 Chemical Master Equation

The Chemical Master Equation (CME) [8] has complete information about time evolution of probability of system's state

$$\frac{\partial \mathcal{P}(x, t)}{\partial t} = \sum_{r=1}^M a_r(x - v_r) \mathcal{P}(x - v_r, t) - a_0(x) \mathcal{P}(x, t). \tag{2.1}$$

Let Q^i be the total possible number of molecules of species S^i . The total number of all possible states of the system is:

$$Q = \prod_{i=1}^N (Q^i + 1).$$

We denote by $\mathcal{I}(x)$ the state-space index of state $x = [X^1, \dots, X^N]$

$$\begin{aligned} \mathcal{I}(x) &= (Q^{N-1} + 1) \cdots (Q^1 + 1) \cdot X^N + \cdots \\ &+ (Q^2 + 1) (Q^1 + 1) \cdot X^3 + (Q^1 + 1) \cdot X^2 + X^1 + 1. \end{aligned}$$

One firing of reaction R_r changes the state from x to $\bar{x} = x - v_r$. The corresponding change in state space index is:

$$\begin{aligned} \mathcal{I}(x) - \mathcal{I}(x - v_r) &= d_r, \\ d_r &= (Q^{N-1} + 1) \cdots (Q^1 + 1) \cdot v_r^N + \cdots \\ &+ (Q^2 + 1) (Q^1 + 1) \cdot v_r^3 + (Q^1 + 1) \cdot v_r^2 + v_r^1. \end{aligned} \tag{2.2}$$

The discrete solutions of the CME (2.1) are vectors in the discrete state space, $\mathcal{P}(t) \in \mathbb{R}^Q$. Consider the diagonal matrix $A_0 \in \mathbb{R}^{Q \times Q}$ and the Toeplitz matrices $A_1, \dots, A_M \in \mathbb{R}^{Q \times Q}$ [14]

$$(A_0)_{i,j} = \begin{cases} -a_0(x_j) & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad (A_r)_{i,j} = \begin{cases} a_r(x_j) & \text{if } i - j = d_r, \\ 0 & \text{if } i - j \neq d_r, \end{cases}$$

as well as their sum $A \in \mathbb{R}^{Q \times Q}$ with entries

$$A = A_0 + A_1 + \cdots + A_M, \quad A_{i,j} = \begin{cases} -a_0(x_j) & \text{if } i = j, \\ a_r(x_j) & \text{if } i - j = d_r, \quad r = 1, \dots, M, \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

where x_j denotes the unique state with state space index $j = \mathcal{I}(x_j)$. In fact matrix A is a square ($Q \times Q$) matrix which contains all the propensity values for each possible value of all species or let's say all possible states of reaction system. All possible states for a reaction system consists of N species where each specie has at most Q^i $i = 1, 2, \dots, N$ value.

The CME (2.1) is a linear ODE on the discrete state space

$$\mathcal{P}' = A\mathcal{P}, \quad \mathcal{P}(\bar{t}) = \delta_{\mathcal{I}(\bar{x})}, \quad t \geq \bar{t}, \tag{2.4}$$

where the system is initially in the known state $x(0) = \bar{x}$ and therefore the initial probability distribution vector $\mathcal{P}(0) \in \mathbb{R}^Q$ is equal to one at $\mathcal{I}(\bar{x})$ and is zero everywhere else. The exact solution of the linear ODE (2.4) is follows:

$$\mathcal{P}(\bar{t} + T) = \exp(TA)\mathcal{P}(\bar{t}) = \exp\left(T \sum_{r=0}^M A_r\right)\mathcal{P}(\bar{t}). \tag{2.5}$$

2.2 Approximation to chemical master equation

Although the CME (2.1) fully describes the evolution of probabilities it is difficult to solve in practice due to large state space. Sandu [14] considers the following approximation of the CME:

$$\frac{\partial \mathcal{P}(x, t)}{\partial t} = \sum_{r=1}^M a_r(\bar{x}) \mathcal{P}(x - v_r, t) - a_0(\bar{x}) \mathcal{P}(x, t), \tag{2.6}$$

where the arguments of all propensity functions have been changed from x or $x - v_j$ to \bar{x} . In order to obtain an exponential solution to (2.6) in probability space we consider the diagonal matrix $\bar{A}_0 \in \mathbb{R}^{Q \times Q}$ and the Toeplitz matrices $\bar{A}_1, \dots, \bar{A}_M \in \mathbb{R}^{Q \times Q}$ [14]. \bar{A}_r matrices are square ($Q \times Q$) matrices are built upon the current state of system in reaction system which is against A_r matrices that contain all possible states of reaction system.

$$(\bar{A}_0)_{i,j} = \begin{cases} -a_0(\bar{x}) & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad (\bar{A}_r)_{i,j} = \begin{cases} a_r(\bar{x}) & \text{if } i - j = d_r, \\ 0 & \text{if } i - j \neq d_r, \end{cases} \tag{2.7}$$

together with their sum $\bar{A} = \bar{A}_0 + \dots + \bar{A}_M$. The approximate CME (2.6) can be written as the linear ODE

$$\mathcal{P}' = \bar{A}\mathcal{P}, \quad \mathcal{P}(\bar{t}) = \delta_{\mathcal{I}(\bar{x})}, \quad t \geq \bar{t},$$

and has an exact solution

$$\mathcal{P}(\bar{t} + T) = \exp(T\bar{A})\mathcal{P}(\bar{t}) = \exp\left(T \sum_{r=0}^M \bar{A}_r\right)\mathcal{P}(\bar{t}). \tag{2.8}$$

2.3 Tau-leaping method

In tau-leap method the number of times a reaction fires is a random variable from a Poisson distribution with parameter $a_r(\bar{x})\tau$. Since each reaction fires independently, the probability that each reaction R_r fires exactly k_r times, $r = 1, 2, \dots, M$, is the product of M Poisson probabilities.

$$\mathcal{P}(K_1=k_1, \dots, K_M=k_M) = \prod_{r=1}^M e^{-a_r(\bar{x})\tau} \frac{(a_r(\bar{x})\tau)^{k_r}}{K_r!} = e^{-a_0(\bar{x})\tau} \prod_{r=1}^M \frac{(a_r(\bar{x})\tau)^{k_r}}{K_r!}.$$

The state vector after these reactions changes as follows:

$$X(\bar{t} + \tau) = \bar{x} + \sum_{r=1}^M K_r v_r. \tag{2.9}$$

The probability to go from state \bar{x} at \bar{t} to state x at $\bar{t} + \tau$, $\mathcal{P}(X(\bar{t} + \tau) = x)$, is the sum of all possible firing reactions which is:

$$\mathcal{P}(X, \bar{t} + \tau) = e^{-a_0(\bar{x})T} \cdot \sum_{k \in \mathcal{K}(x-\xi)} \prod_{r=1}^M \frac{(a_r(\bar{x}T))^{k_r}}{K_r!}.$$

Equation (2.8) can be approximated by product of each matrix exponential:

$$\mathcal{P}(\bar{t} + T) = \exp(T\bar{A}_0) \exp(T\bar{A}_1) \exp(T\bar{A}_r) \mathcal{P}(\bar{t}). \tag{2.10}$$

It has been shown in [14] that the probability given by the tau-leaping method is exactly the probability evolved by the approximate solution (2.10).

3 Approximations to the Exponential Solution

3.1 Strang splitting

In order to improve the approximation of the matrix exponential in (2.10) we consider the symmetric Strang splitting [15]. For $T = n\tau$ Strang splitting applied to an interval of length τ leads to the approximation

$$\mathcal{P}(\bar{t} + i\tau) = e^{\tau/2\bar{A}_r} e^{\tau/2\bar{A}_1} e^{\tau\bar{A}_0} e^{\tau/2\bar{A}_1} e^{\tau/2\bar{A}_r} \mathcal{P}(\bar{t} + (i-1)\tau), \tag{3.1}$$

where the matrices \bar{A}_r are defined in (2.7).

3.2 Column based splitting

In column based splitting the matrix A (2.3) is decomposed in a sum of columns

$$A = \sum_{j=1}^Q A_j, \quad A_j = c_j e_j^T.$$

Each matrix A_j has the same j -th column as the matrix A , and is zero everywhere else. Here c_j is the j th column of matrix A and e_j is the canonical vector which is zero every where except the j th component. The exponential of τA_j is:

$$e^{\tau A_j} = \sum_{k \geq 0} \frac{\tau^k (A_j)^k}{k!}. \tag{3.2}$$

Since $e_j^T c_j$ is equal to the j -th diagonal entry of matrix A :

$$e_j^T c_j = -a_0(x_j),$$

the matrix power A_j^k reads

$$A_j^k = c_j e_j^T c_j e_j^T \cdots c_j e_j^T = (-a_0(x_j))^{k-1} c_j e_j^T = (-a_0(x_j))^{k-1} A_j.$$

Consequently the matrix exponential (3.2) becomes

$$e^{\tau A_j} = I + \sum_{k \geq 1} \frac{(-\tau a_0(x_j))^{k-1}}{k!} (\tau A_j) = I + S_j \tau A_j, \quad S_j = \sum_{k \geq 1} \frac{(-\tau a_0(x_j))^{k-1}}{k!}.$$

We have

$$e^{\tau A} = e^{\tau \sum_{j=1}^Q A_j} \approx \prod_{j=1}^Q e^{\tau A_j} \approx \prod_{j=1}^Q (I + S_j \tau A_j)$$

and the approximation to the CME solution reads

$$\mathcal{P}(\bar{t} + i\tau) \approx \prod_{j=1}^Q (I + S_j \tau A_j) P(\bar{t} + (i - 1)\tau).$$

3.3 Accelerated tau-leaping

In this approximation method we build the matrices

$$(B_r)_{i,j} = \begin{cases} -a_r(x_j) & \text{if } i = j, \\ a_r(x_j) & \text{if } i - j = d_r, \\ 0 & \text{otherwise,} \end{cases}$$

where $a_r(x)$ are the propensity functions. The matrix A in (2.3) can be written as

$$A = \sum_{r=1}^M B_r.$$

The solution of the linear CME (2.5) can be approximated by

$$\mathcal{P}(\bar{t} + \tau) = e^{\tau A} \mathcal{P}(\bar{t}) \approx e^{\tau B_1} e^{\tau B_2} \dots e^{\tau B_M} P(\bar{t}). \tag{3.3}$$

Note that the evolution of state probability by $e^{\tau B_j} \cdot P(\bar{t})$ describes the change in probability when only reaction j fires in the time interval τ . The corresponding evolution of the number of molecules that samples the evolved probability is

$$x(\bar{t} + \tau) = x(\bar{t}) + V_j K(a_j(x(\bar{t}))\tau),$$

where $K(a_j(x(\bar{t}))\tau)$ is a random number drawn from a Poisson distribution with parameter $a_j(x(\bar{t}))\tau$, and V_j is the j -th column of stoichiometry matrix.

The approximate solution (3.3) accounts for the change in probability due to a sequential firing of reactions $M, M - 1, \dots, 1$. Sampling from the resulting probability density can be done by changing the system state sequentially consistent with the firing of each reaction. This leads to algorithm (3.4). The accelerated tau-leap method (3.4) provides accurate and stable results especially for large time steps since the change in state is accomplished through a sequential firing of reactions. From the complexity point of view, the number of propensity function calculations and Poisson random numbers generated are the same as for traditional tau-leap. We note that in traditional tau-leap a vector of Poisson random variables is generated at once, which is possibly more efficient than generating the random numbers one at a time.

$$\begin{aligned} \hat{X}_M &= x(\bar{t}) \\ \text{for } i &= M, M - 1, \dots, 1 \\ \hat{X}_{i-1} &= \hat{X}_i + V_i K(a_i(\hat{X}_i)\tau) \\ x(\bar{t} + \tau) &= \hat{X}_0. \end{aligned} \tag{3.4}$$

Moreover, (3.3) can also be written as:

$$\begin{aligned} \mathcal{P}(\bar{t} + \tau) &\approx e^{\tau B_1} e^{\tau B_2} \dots e^{\tau B_M} P(\bar{t}) \\ &\approx (e^{\tau B_1} e^{\tau B_2} \dots e^{\tau B_{\frac{M}{2}-1}}) (e^{\tau B_{\frac{M}{2}}} e^{\tau B_{\frac{M}{2}+1}} \dots e^{\tau B_M}) P(\bar{t}) \end{aligned} \tag{3.5}$$

Then, (3.4) can be written as:

$$\begin{aligned} \hat{X}_M &= x(\bar{t}) \\ \text{for } i &= M, M-1, \dots, \frac{M}{2} \\ \hat{X}_{i-1} &= \hat{X}_i + V_i K(a(\hat{X}_i)\tau) \\ \text{for } i &= \frac{M}{2} - 1, \dots, 1 \\ \hat{X}_{i-1} &= \hat{X}_i + V_i K(a(\hat{X}_{\frac{M}{2}-1})\tau) \\ x(\bar{t} + \tau) &= \hat{X}_0. \end{aligned} \tag{3.6}$$

3.4 Symmetric accelerated tau-leaping

A more accurate version of accelerated tau-leaping can be constructed by using symmetric Strang splitting (3.1) to approximate the matrix exponential in (3.3). Following the procedure used to derive (3.4) leads to the following sampling algorithm:

$$\begin{aligned} \hat{X}_M &= x(\bar{t}) \\ \text{for } i &= M, M-1, \dots, 1 \\ \hat{X}_{i-1} &= \hat{X}_i + V_i K(a_i(\hat{X}_i)\tau/2) \\ \text{for } i &= 1, 2, \dots, M \\ \hat{X}_i &= \hat{X}_i + V_{i-1} K(a_i(\hat{X}_{i-1})\tau/2) \\ x(\bar{t} + \tau) &= \hat{X}_M. \end{aligned} \tag{3.7}$$

The symmetric accelerated tau-leap algorithm (3.7) is twice as expensive as the accelerated tau-leap (3.6) for the same value of τ since it computes the propensity functions and generates Poisson random variables two times per step.

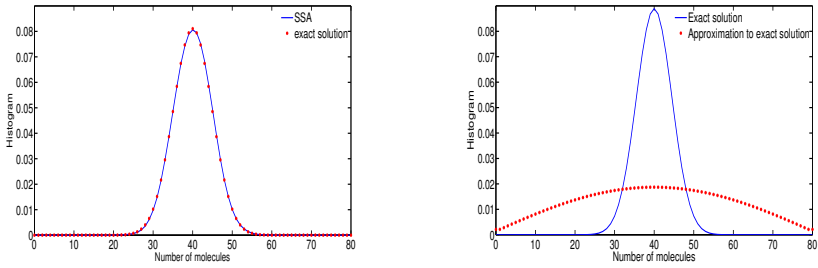
4 Numerical Experiments

The above approximation techniques are used to solve two test systems, reversible isomer and the Schlogl reactions [5]. The experimental results are presented in following sections.

4.1 Isomer reaction

The reversible isomer reaction system is [5]





(a) 10,000 SSA runs versus the exact solution (2.5) (b) Exact solution (2.5) versus the approximation to exact solution using sum of exponentials (2.8)

Figure 1. Histograms of the isomer system (4.1) results at the final time $T = 10$.

The stoichiometry matrix and the propensity functions are:

$$V = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \begin{aligned} a_1(x) &= c_1 x_1, \\ a_2(x) &= c_2 x_2. \end{aligned}$$

The reaction rate values are $c_1 = 10$, $c_2 = 10$ (units), the time interval is $[0, T]$ with $T = 10$ (time units), initial conditions are $x_1(0) = 40$, $x_2(0) = 40$ molecules, and maximum values of species are $Q^1 = 80$ and $Q^2 = 80$ molecules.

The exact exponential solution of CME obtained from (2.5) is a joint probability distribution vector for the two species at final time. Figure 1a shows that the histogram of 10,000 SSA solutions is very close to the exact exponential solution. The approximate solution using the sum of exponentials (2.8) is illustrated in Figure 1b. This approximation is not very accurate since it uses only the current state of the system. Other approximation methods based on the product of exponentials (2.10), Strang splitting, (3.1) and column based splitting are not very accurate, and consequently we choose not to report their results. The reason of poor approximation of product of exponentials and Strang splitting methods is due to the error propagation that occurs during successive matrix vector multiplications.

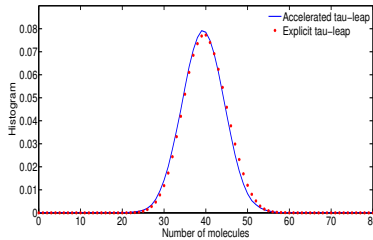


Figure 2. Isomer system (4.1) solutions provided by the traditional tau-leap (2.9) and by accelerated tau-leap (3.4) methods at the final time $T = 10$ (units). A small time step of $\tau = 0.01$ (units) is used. The number of samples for both methods is 10,000.

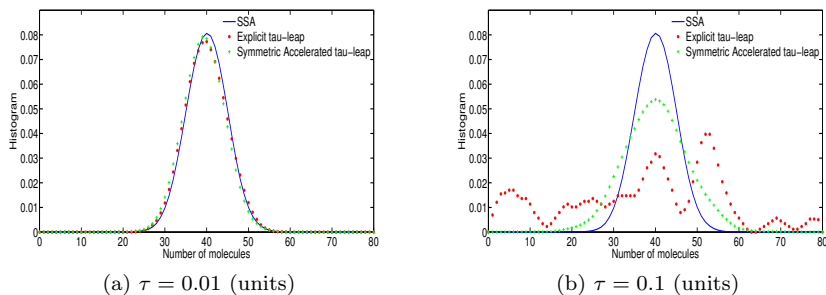


Figure 3. Histograms of isomer system (4.1) solutions obtained with SSA, traditional tau-leap (2.9), and symmetric accelerated tau-leap (3.7) methods at the final time $T = 10$. The number of samples is 10,000 for all methods.

The results reported in Figure 2 indicate that for small time steps τ the accelerated tau-leap (3.4) solution is very close to the results provided by traditional explicit tau-leap.

Symmetric accelerated tau-leap method (3.7) yields even better results, as shown in Figure 3. For small time steps the traditional and symmetric accelerated methods give similar results, however, for large time steps, the results of the symmetric accelerated method is considerably more stable.

To assess accuracy we measure the numerical error as the the difference between the PDF of each algorithm and the PDF of SSA:

$$error = \frac{\sum_{i=0}^{Q(1)} |SSA_{molecules}(i) - NewAlgorithm_{molecules}(i)|}{Q(1)}.$$

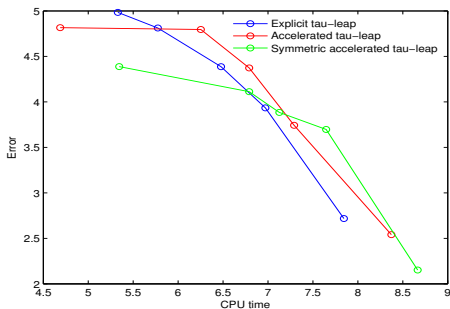


Figure 4. Plot of errors vs. CPU times for different algorithms. The points on the graphs correspond to the following five time steps: $\tau = 0.01$, $\tau = 0.05$, $\tau = 0.075$, $\tau = 0.1$, and $\tau = 0.15$ (units). The final time $T = 10$ (units). Each point is obtained from 10,000 samples/runs.

Figure 4 plots the errors given by different algorithms versus computational time. Each point on the curves corresponds to a different time step τ ; specifically, 10,000 simulations with that value of τ are carried out and the samples

are used to obtain a histogram that approximates the PDF. For larger time steps τ (smaller CPU times) the errors of symmetric accelerated tau-leap and accelerated tau-leap are smaller than the error of traditional tau-leap. The two new algorithms are more effective for lower accuracy computations. For small time steps τ (target CPU times) the traditional tau-leap is the most efficient. The CPU time for the symmetric accelerated tau-leap the largest among the three methods for a given time step since it requires computing propensity functions twice per step.

4.1.1 Absolute stability and stiffness

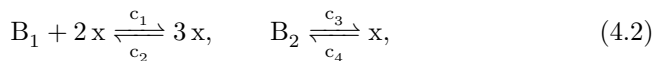
To assess the stability of the new algorithms we apply them to solve the isomer reaction [5]. It turns out that for this test problem the mean and variance of the solution obtained by accelerated tau-leap and by symmetric accelerated tau-leap are the same as for traditional tau-leap, since the only change in these algorithms is the sequential firing of reactions. The procedure outlined in [5] for obtaining the absolute stability is to fix τ and let the number of steps n tend to infinity. Having $X^* = X^{\text{theoretical}}(\infty) = X_1^{\text{theoretical}}(\infty)$ and $\lambda = c_1 + c_2$ in isomer reaction, the following asymptotic values of the mean and variance are obtained [5]

$$E[X(\infty)] = E(X^*), \quad \text{Var}[X(\infty)] = \frac{2}{2 - \lambda\tau} \text{Var}(X^*).$$

The mean value given by tau-leap method converges to the theoretical mean value, while the variance does not. The same conclusions hold for the accelerated tau-leap and symmetric accelerated tau-leap methods.

4.2 Schlogl reaction

We next consider the Schlogl reaction system [5]



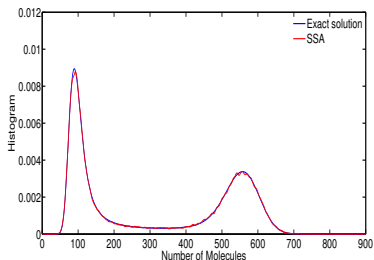
whose solution has a bi-stable distribution. Let N_1, N_2 be the numbers of molecules of species B_1 and B_2 , respectively. The reaction stoichiometry matrix and the propensity functions are:

$$\begin{aligned} V &= [1 \quad -1 \quad 1 \quad -1], \\ a_1(x) &= \frac{c_1}{2} N_1 x(x - 1), \quad a_2(x) = \frac{c_2}{6} N_1 x(x - 1)(x - 2), \\ a_3(x) &= c_3 N_2, \quad a_4(x) = c_4 x. \end{aligned}$$

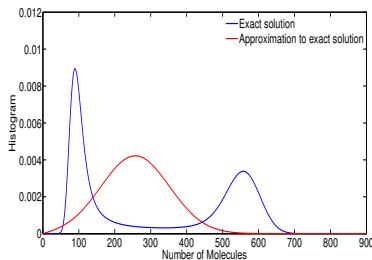
The following parameter values (each in appropriate units) are used:

$$\begin{aligned} c_1 &= 3 \times 10^{-7}, \quad c_2 = 10^{-4}, \quad c_3 = 10^{-3}, \\ c_4 &= 3.5, \quad N_1 = 1 \times 10^5, \quad N_2 = 2 \times 10^5 \end{aligned}$$

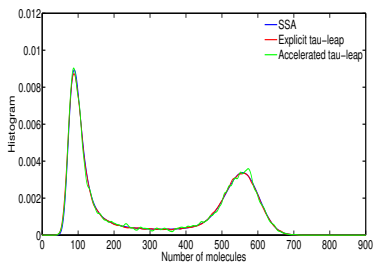
with the final time $T = 4$ (units), the initial condition $x(0) = 250$ molecules, and the maximum values of species $Q^1 = 900$ molecules.



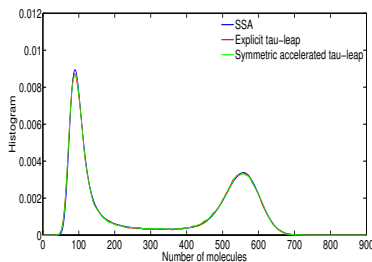
(a) 10,000 SSA runs versus the exact solution (2.5)



(b) Exact solution (2.5) versus the approximation to exact solution using sum of exponentials (2.8)

Figure 5. Histograms of Schlogl system (4.2) results at final time $T = 4$ (units).

(a) Traditional tau-leap (2.9) and accelerated tau-leap (3.4)



(b) Traditional tau-leap (2.9) and symmetric accelerated tau-leap (3.7)

Figure 6. Histograms of Schlogl system (4.2) solutions with $\tau = 0.0001$ (units), final time $T = 4$ (units), and 10,000 samples.

Figure 5a illustrates the result of exact exponential solution (2.5) versus SSA. Figure 5b reports the sum of exponentials (2.8) result which is not a very good approximation. The product of exponentials (2.10) and Strang splitting (3.1) results are not reported here since they provide inaccurate approximations.

Figures 6a and 6b present the results obtained with the accelerated tau-leap and the symmetric tau-leap, respectively. For small time step the results are very accurate. For large step sizes the results become less accurate but continue to be more stable than tau-leap. Systems having a large number of reactions may be more affected by the low accuracy, and improvements such as the ones described in equations (3.5) and (3.6) may prove helpful.

5 Conclusions

This study proposes new numerical solvers for stochastic simulations of chemical kinetics. The proposed approach exploits the linearity of the CME and the exponential form of its exact solution. The matrix exponential appearing in

the CME solution is approximated as a product of simpler matrix exponentials. This leads to an approximate (“numerical”) solution of the probability density evolved to a future time. The solution algorithms sample exactly this approximate probability density and provide extensions of the traditional tau-leap approach.

Different approximations of the matrix exponential lead to different numerical algorithms: Strang splitting, column splitting, accelerated tau-leap, and symmetric accelerated tau-leap. Numerical results illustrate that the new approximation methods are more stable than explicit tau-leap especially for large time steps, but are less accurate for some reaction systems. Despite this fact the class novel numerical solvers proposed herein is of interest since it is based on a totally different approach than the one used to derive classical schemes. Specifically, one first discretizes the chemical master equation, then draws samples from the resulting probability density. Current work by the authors focuses on improving the accuracy of the new family of approximation techniques.

Acknowledgements

This work was partially supported by awards NSF DMS-1419003, NSF CCF-1218454, AFOSR FA9550-12-1-0293-DEF, AFOSR 12-2640-06, and by the Computational Science Laboratory at Virginia Tech.

Appendix A. Example

We exemplify the process of building matrix A (2.3) for the Schlogl and isomer reactions.

Isomer reaction Here for simplicity, we exemplify the implementation of the system for the maximum values of species $Q^1 = 2$ and $Q^2 = 2$. According to (2.1), $Q = (Q^1 + 1) \times (Q^2 + 1) = 3^2$.

The vector d according to (2.2) is $[2, -2]$. The state matrix which contains all possible states has dimension $81^2 \times 2$ matrix:

$$\mathbf{x} = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \end{bmatrix}^\top \in \mathbb{R}^{3^2 \times 2}.$$

The matrix $\mathbf{A} \in \mathbb{R}^{Q \cdot Q \times Q \cdot Q}$ As an example for a maximum number of species

$Q^1 = 2, Q^2 = 2$ the matrix \mathbf{A} is:

$$\mathbf{A} = \begin{bmatrix} -a_0(\mathbf{x}_{1,:}) & 0 & a_2(\mathbf{x}_{3,:}) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -a_0(\mathbf{x}_{2,:}) & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ a_1(\mathbf{x}_{1,:}) & 0 & -a_0(\mathbf{x}_{3,:}) & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & a_1(\mathbf{x}_{2,:}) & 0 & \ddots & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & a_1(\mathbf{x}_{3,:}) & 0 & \ddots & 0 & a_2(\mathbf{x}_{7,:}) & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & \ddots & 0 & a_2(\mathbf{x}_{8,:}) & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & -a_0(\mathbf{x}_{7,:}) & 0 & a_2(\mathbf{x}_{9,:}) \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & -a_0(\mathbf{x}_{8,:}) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_1(\mathbf{x}_{7,:}) & 0 & -a_0(\mathbf{x}_{9,:}) \end{bmatrix} \in \mathbb{R}^{9 \times 9}.$$

Schlogl reaction Here for simplicity, we exemplify the implementation of the system for the maximum value of the number of molecules $Q^1 = 5$. According to (2.1) the dimensions of A are: $(Q^1 + 1 \times Q^1 + 1) = 6 \times 6$. The vector d (2.2) for this system $[1, -1, 1, -1]$. All possible states for this system are contained in the state vector

$$\mathbf{x} = [0, 1, 2, \dots, 5]^T \in \mathbb{R}^{1 \times 6}.$$

As an example matrix A for maximum number of molecules $Q = 5$ is the following tridiagonal matrix:

$$\mathbf{A} = \begin{bmatrix} -a_0(\mathbf{x}_1) & a_2(\mathbf{x}_2) + a_4(\mathbf{x}_2) & 0 & 0 & 0 & 0 \\ a_1(\mathbf{x}_1) + a_3(\mathbf{x}_1) & -a_0(\mathbf{x}_2) & \ddots & 0 & 0 & 0 \\ 0 & a_1(\mathbf{x}_2) + a_3(\mathbf{x}_2) & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \ddots & \ddots & a_2(\mathbf{x}_5) + a_4(\mathbf{x}_5) & 0 \\ 0 & 0 & 0 & \ddots & -a_0(\mathbf{x}_5) & a_2(\mathbf{x}_6) + a_4(\mathbf{x}_6) \\ 0 & 0 & 0 & 0 & a_1(\mathbf{x}_5) + a_3(\mathbf{x}_5) & -a_0(\mathbf{x}_6) \end{bmatrix} \in \mathbb{R}^{6 \times 6}.$$

References

- [1] T.H. Ahn and A. Sandu. Implicit simulation methods for stochastic chemical kinetics, 2013. <http://arxiv.org/abs/1303.3614>
- [2] T.H. Ahn, A. Sandu, L. Watson, C. Shaffer, Y. Cao and W. Baumann. Parallel load balancing strategies for ensembles of stochastic biochemical simulations. *Tech. rep., Virginia Tech.*, 2012.
- [3] Y. Cao, D.T. Gillespie and L.R. Petzold. The slow-scale stochastic simulation algorithm. *J. Chemical Physics*, **122**:014116, 2005. <http://dx.doi.org/10.1063/1.1824902>.
- [4] Y. Cao, H. Li and L. Petzold. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chemical Physics*, **9**(121):4059–4067, 2004. <http://dx.doi.org/10.1063/1.1778376>.

- [5] Y. Cao, R. Petzold, M. Rathinam and D. Gillespie. The numerical stability of leaping methods for stochastic simulation of chemically reacting systems. *J. Chemical Physics*, **121**(24):12169–12178, 2004. <http://dx.doi.org/10.1063/1.1823412>.
- [6] S. Engblom. Numerical methods for the chemical master equation. *Ph.D. thesis, Uppsala University, Department of Information Technology*, 2006.
- [7] D. Gillespie and L. Petzold. Improved leap-size selection for accelerated stochastic simulation. *J. Chemical Physics*, **119**(16):8229–8234, 2003. <http://dx.doi.org/10.1063/1.1613254>.
- [8] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Chemical Physics*, **81**(25):2340–2361, 1977. <http://dx.doi.org/10.1021/j100540a008>.
- [9] D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chemical Physics*, **115**(4):1716–1733, 2001. <http://dx.doi.org/10.1063/1.1378322>.
- [10] T. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *J. Chemical Physics*, **57**(7):2976–2978, 1972. <http://dx.doi.org/10.1063/1.1678692>.
- [11] S. MacNamara, K. Burrage and R.B. Sidje. Multiscale modeling of chemical kinetics via the master equation. *SIAM J. Multiscale Modeling and Simulation*, **4**(6):1146–1168, 2008. <http://dx.doi.org/10.1137/060678154>.
- [12] M. Rathinam, L. Petzold, Y. Cao and D. Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *J. Chemical Physics*, **119**(24):12784, 2003. <http://dx.doi.org/10.1063/1.1627296>.
- [13] M. Rathinam, L. Petzold, Y. Cao and D. Gillespie. Consistency and stability of tau leaping schemes for chemical reaction systems. *SIAM J. Multiscale Modeling and Simulation*, **3**(4):867–895, 2005. <http://dx.doi.org/10.1137/040603206>.
- [14] A. Sandu. A new look at chemical master equation. *Numer. Algorithms*, **65**(3):485–498, 2013. <http://dx.doi.org/10.1007/s11075-013-9758-z>.
- [15] G. Strang. On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, **3**(5):506–517, 1968. <http://dx.doi.org/10.1137/0705041>.