

Efficient Leaping Methods for Stochastic Chemical Systems

Ioana Cipcigan* Muruhan Rathinam*

November 18, 2008

Abstract. Well stirred chemical reaction systems which involve small numbers of molecules for some species have a stochastic behavior and can be modeled by a continuous time, discrete state Markov process. An exact method for simulating the time evolution of the system is the Stochastic Simulation Algorithm, but this method is extremely slow for realistic biological systems. We propose an adaptive leaping method based on local error formulas and we compare the adaptive method with the exact method and with the explicit tau leaping method with constant time steps.

Key words. stochastic chemical kinetics, stochastic simulation algorithm, tau-leaping

AMS subject classifications (2000). 60H35, 65C30

1 Introduction

Given a mixture of N molecular species S_1, \dots, S_N that chemically interact inside some fixed volume Ω at a constant temperature T , through M elementary chemical reactions channels R_1, \dots, R_M , we are interested in the behavior of species population variables.

Chemical systems that involve small numbers of molecules for some species cannot be adequately modeled by a continuous and deterministic model. In this case, the behavior of the system is rather discrete and stochastic. Well stirred chemical reaction systems can be modeled by a continuous time, discrete state Markov process, which has the state $X(t) \in \mathbb{Z}_+^N$, where $X_i(t)$ ($i = 1, \dots, N$) is the number of molecules of species S_i at time t . Associated with each reaction R_j ($j = 1, \dots, M$) is a propensity function $a_j(x) : \mathbb{Z}_+^N \rightarrow \mathbb{R}_+$, where $a_j(x)dt$ is the probability, given $X(t) = x$, that one R_j reaction will occur in the next infinitesimal time interval $[t, t + dt)$. The dynamic of the system is completely characterized by the propensities

*Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, {ioana1,muruhan}@umbc.edu .

functions and the state-change vectors ν_j , where ν_{ij} represents the change in the number of S_i molecules produced by one R_j reaction.

The time evolution of the probability density function of $X(t)$, $P(x, t|x_0, t_0)$, the probability that $X(t) = x$, given $X(t_0) = x_0$ is characterized by the Kolmogorov Forward Equation $\frac{\partial P}{\partial t} = PA$, where A is the transition rate matrix [10]. This leads to the chemical master equation, a differential equation for time evolution of probabilities. But for most realistic chemical systems, this is hard to solve, analytically or numerically. An equivalent way of solving this problem is to generate realizations of sample trajectories that evolve according to the same probability. This can be done by stochastic simulation algorithm (SSA), which was proposed by Gillespie in [2]. However, even this method is extremely slow for realistic biological systems, since it involves simulating one reaction at a time.

An approximate method, the *tau leaping method*, has been proposed in [4] to accelerate the SSA. The basic idea is to approximate the number of times each reaction channel R_j fires during the time interval time $(t, t + \tau]$ by a Poisson random variable with mean $a_j(x)\tau$, assuming that the time step τ is small enough such that the propensity functions are almost constant.

This gives the *Explicit Tau Method*:

$$X^{(et)}(t + \tau) = x + \sum_{j=1}^M \nu_j P_j(a_j(x)\tau).$$

Since the Poisson distribution with nonzero mean assigns nonzero probability to arbitrarily large numbers, the method can produce negative states, while the exact method does not. The following bounding algorithm has been proposed in [9] to avoid this problem:

Suppose that the state x_n reached after leaping is a negative state, and the number of times reaction R_j fires during this leap is K_j , for $j = 1, \dots, M$. Then in order to obtain a nonnegative state, while x_n has negative components, decrease by one each K_j and update x_n accordingly.

An adaptive method for selecting the time step using the local error formulas is presented and a comparison of this method with the constant time step *Explicit Tau Method*.

Finally, a deterministic versus stochastic approach is presented and the connection with stochastic differential equations is discussed.

2 Adaptive leaping method

We believe that an efficient leaping method should include a strategy for adaptive step size selection. Based on local error formulas we develop an adaptive leaping method for this problem. The local error formulas for the mean and covariance of explicit tau have been derived in [9]. Here we explain the derivation briefly.

Let the multi-index $k = (k_1, \dots, k_l)$ denote a sequence of reactions events R_{k_j} happening in this order. Define $p(k; x, \tau)$ to be the probability that the sequence of reactions that occurred in $(t, t + \tau]$ is precisely k , given that $X(t) = x$. Then we have:

$$p((\cdot); x, \tau) = e^{-a_0(x)\tau}, a_0(x) = \sum_{j=1}^M a_j(x),$$

$$p((k_1, \dots, k_l); x, \tau) = \int_0^\tau p((k_1, \dots, k_{l-1}; x, s) a_{k_l}(x + \nu_{k_1} + \dots + \nu_{k_{l-1}}) e^{-a_0(x + \nu_{k_1} + \dots + \nu_{k_l})(\tau-s)} ds.$$

This implies that $p(k; x, \tau) = O(\tau^{|k|})$. We want to compute $p(k; x, t)$ for terms up to $O(\tau^2)$ for general M and N . Taylor Expansion for the terms up to $|k| = 2$ gives:

$$p((j); x, \tau) = \tau a_j(x) - \frac{1}{2} \tau^2 \sum_{j=1}^M a_j(x) [a_j(x + \nu_j) + a_j(x)] + O(\tau^3), \quad (2.1)$$

$$p((j_1, j_2); x, \tau) = \frac{1}{2} \tau^2 a_{j_1}(x) a_{j_2}(x + \nu_{j_1}) + O(\tau^3). \quad (2.2)$$

The conditional r^{th} moment of the increment $X(t + \tau) - X(t)$ can be computed follows:

$$E[(X(t + \tau) - X(t))^r | X(t) = x] = \sum_{l=1}^{\infty} \sum_{|k|=l} p(k; x, \tau) \left(\sum_{\alpha=1}^l \nu_{k_\alpha} \right)^r. \quad (2.3)$$

Using formulas (2.1)–(2.3), we get the formula for the general r^{th} moment:

$$\begin{aligned} E[(X(t + \tau) - X(t))^r | X(t) = x] &= \tau \sum_{j=1}^M \nu_j^r a_j(x) - \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1}^r a_{j_1}(x) a_{j_2}(x) \\ &\quad - \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1}^r a_{j_1}(x) a_{j_2}(x + \nu_{j_1}) \\ &\quad + \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M (\nu_{j_1} + \nu_{j_2})^r a_{j_1}(x) a_{j_2}(x + \nu_{j_1}) + O(\tau^3). \end{aligned}$$

For the explicit tau method, the conditional r^{th} moment of the increment is given by:

$$E[(X^{(et)}(t + \tau) - X^{(et)}(t))^r | X^{(et)}(t) = x] = \tau \sum_{j=1}^M \nu_j^r a_j(x) + O(\tau^2).$$

This gives the following formulas for the error in mean and the error in covariance:

$$Error_{mean}(\tau) = \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} a_{j_2}(x) [a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)],$$

$$\begin{aligned}
Error_{covariance}(\tau) &= \frac{1}{2}\tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1}^2 a_{j_2}(x) [a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)] \\
&+ \frac{1}{2}\tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} \nu_{j_2} a_{j_1}(x) [a_{j_2}(x + \nu_{j_1}) - a_{j_2}(x)] \\
&+ \frac{1}{2}\tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} \nu_{j_2} a_{j_2}(x) [a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)].
\end{aligned}$$

Following the same ideas as in ODE solvers [11], we use the local error formulas to develop a criteria for adaptively choosing the time step. The criteria for choosing the time step τ is given by the following conditions:

$$\|Error_{mean}(\tau)\| < ATOL + RTOL\|x\|,$$

$$\|Error_{covariance}(\tau)\| < ATOL + RTOL\|x^2\|,$$

$$\tau < \tau_{stable}.$$

The last condition is a stability condition and it comes from the explicit nature of the method, and $ATOL$ and $RTOL$ are given tolerances.

The convergence of explicit tau method has been proved in [9]. The method is first order convergent for all the moments.

3 Numerical Results

For testing the new adaptive method, we chose an oscillatory example. We use an example from biology, the repressilator [5], the first oscillatory synthetic circuit:

$$\begin{aligned}
\dot{r}_a &= -r_a + \frac{50}{10 + c^2}, \\
\dot{a} &= -0.1a + 10r_a, \\
\dot{r}_b &= -r_b + \frac{50}{10 + a^2}, \\
\dot{b} &= -0.1b + 10r_b, \\
\dot{r}_c &= -r_c + \frac{50}{10 + b^2}, \\
\dot{c} &= -0.1c + r_c.
\end{aligned}$$

Here a, b, c represent three types of proteins and r_a, r_b, r_c represent the corresponding mRNAs. This system has 12 reaction channels. For each species there is a decay reaction

and a production reaction. For the proteins this is the translation and for mRNAs this is transcription. The initial conditions are: $r_a(0) = r_b(0) = r_c(0) = a(0) = b(0) = c(0) = 0$.

For the stochastic model, the propensity functions and the state change vectors are given bellow:

$$\begin{aligned}
a_1 &= r_a; & \nu_1 &= (-1, 0, 0, 0, 0, 0); \\
a_2 &= \frac{50}{10 + c^2}; & \nu_2 &= (1, 0, 0, 0, 0, 0); \\
a_3 &= 0.1a; & \nu_3 &= (0, -1, 0, 0, 0, 0); \\
a_4 &= 10r_a; & \nu_4 &= (0, 1, 0, 0, 0, 0); \\
a_5 &= r_b; & \nu_5 &= (0, 0, -1, 0, 0, 0); \\
a_6 &= \frac{50}{10 + a^2}; & \nu_6 &= (0, 0, 1, 0, 0, 0); \\
a_7 &= 0.1b; & \nu_7 &= (0, 0, 0, -1, 0, 0); \\
a_8 &= 10r_b; & \nu_8 &= (0, 0, 0, 1, 0, 0); \\
a_9 &= r_c; & \nu_9 &= (0, 0, 0, 0, -1, 0); \\
a_{10} &= \frac{50}{10 + b^2}; & \nu_{10} &= (0, 0, 0, 0, 1, 0); \\
a_{11} &= 0.1c; & \nu_{11} &= (0, 0, 0, 0, 0, -1); \\
a_{12} &= r_c; & \nu_{12} &= (0, 0, 0, 0, 0, 1).
\end{aligned}$$

Firstly, to show that the stochastic model differs from the deterministic one, we solve the deterministic problem using Matlab's `ode45` solver. Figure 1(a) plots the deterministic solution obtained. Then we solve this problem using the exact method SSA. Figure 1(b) shows three sample trajectories generated by this method. This shows a big difference between the deterministic model and the stochastic one. Figure 1(c) shows the mean trajectory computed by SSA, and again we can see the difference between this and the deterministic solution. The algorithm for generating Poisson random variables is given in [7]

Further, we want to compare the adaptive time step method with the explicit tau method with constant time step. The constant time step for the explicit tau is chosen to be the average time step taken by the adaptive method. We compare these methods with the exact method, SSA. We first look at the histograms of the final state produced by each of them for one species, the protein b . These are plotted in Figure 2 and show a similar behavior.

We also compare the mean and the standard deviation for all the species, at the final time $t = 100$ and we obtain similar results. The results are shown in Figure 3.

To show that the adaptive time step method captures better the fluctuations in the system we use another comparison criterion, the Kolmogorov distance. The Kolmogorov distance is defined by $d(F_X, F_Y) = \sup_{x \in \mathbb{Z}} |F_X(x) - F_Y(x)|$, where F_X is the distribution function of random variable X and F_Y is the distribution function of random variable Y . Using this criterion we compute $d(\hat{F}_{SSA}, \hat{F}_{AET})$ and $d(\hat{F}_{SSA}, \hat{F}_{ET})$ for each species, where \hat{F}_{SSA} is the

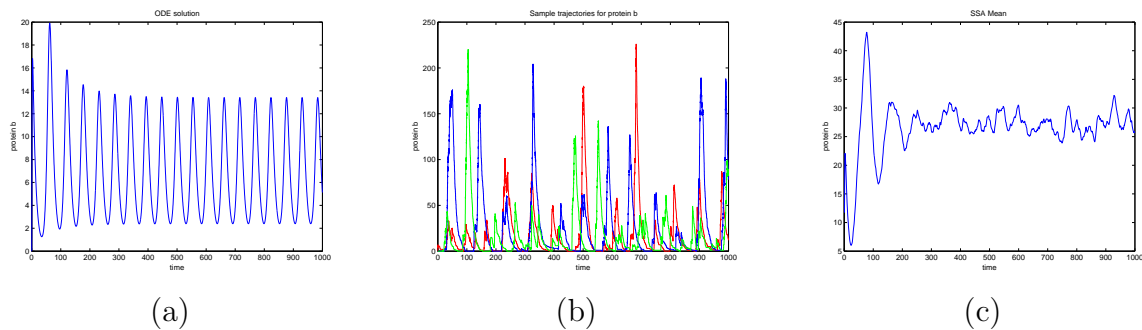


Figure 1: (a) ODE solution for protein b . (b) Sample trajectories for protein b simulated by SSA. (c) Mean trajectory for protein b simulated by SSA.

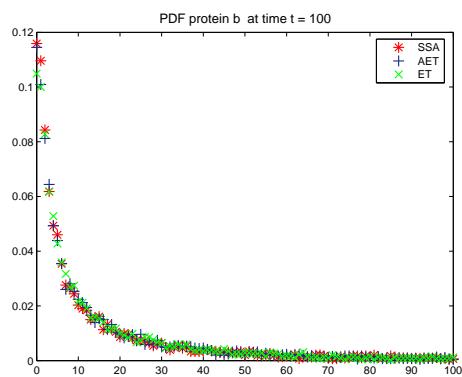


Figure 2: Histogram protein b .

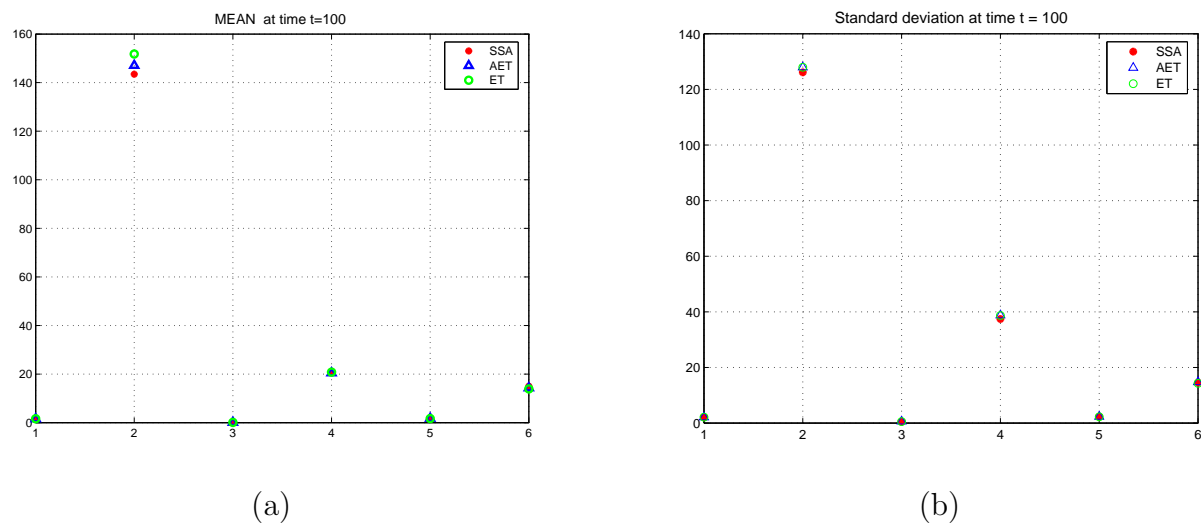


Figure 3: (a) Mean protein b . (b) Standard deviation protein b .

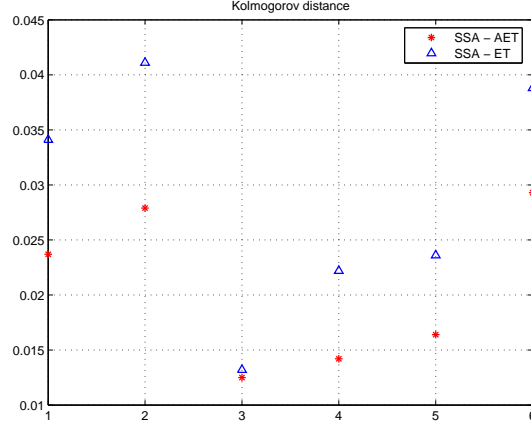


Figure 4: Kolmogorov distance.

MOMENTS	SSA	AET	ET
MEAN	9.4859	9.4334	8.8471
STD	14.5355	14.5896	14.2507

Table 1: Hitting time moments.

numerical estimation of the CDF of SSA, \hat{F}_{AET} is the numerical estimation of the CDF of the adaptive method, and \hat{F}_{ET} is the numerical estimation of the CDF of the explicit tau method. We can see that the adaptive method performs better than the constant time step method. These results are shown in Figure 4.

We also record the first time the number of molecules of protein b reaches or exceeds a critical level, and then we compute the statistics for this “hitting time” as computed by all three methods. Table 1 shows the mean and the standard deviation, and the adaptive method has a better behavior than the constant time step method.

To study the oscillatory behavior of the system, we compared the power spectrum for protein b , computed by these three methods. These results are shown in Figure 5.

4 Deterministic versus stochastic models

When the number of molecules of each species is large enough, such that $a_j(x)\tau \gg 1$, $j = 1, \dots, M$, the Poisson random variable $P(a_j(x), \tau)$ can be approximated by the normal random variable $N(a_j(x)\tau, a_j(x)\tau)$ and this leads to *Langevin method* [4]:

$$X(t + \tau) = x + \tau \sum_{j=1}^M \nu_j a_j(x) + \sqrt{\tau} \sum_{j=1}^M \nu_j \sqrt{a_j(x)} N_j(0, 1)$$

and implies solving a stochastic differential equation(SDE):

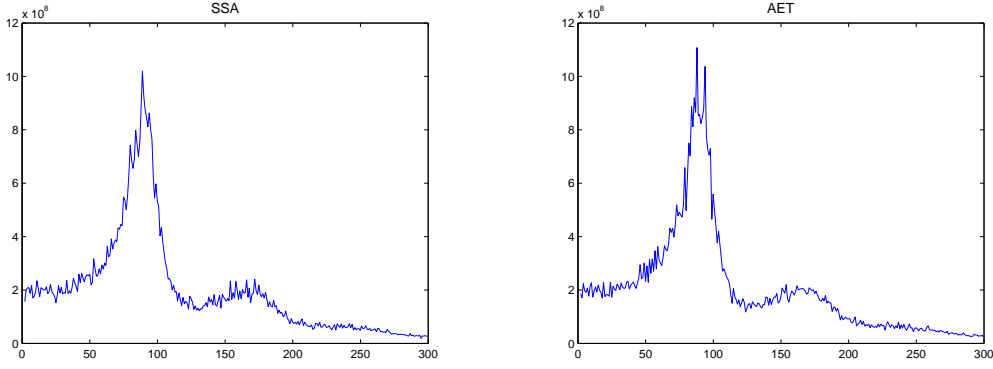


Figure 5: (a) Power spectrum protein b : SSA. (b) Power spectrum protein b : AET.

$$\frac{dX(t)}{dt} = \sum_{j=1}^M \nu_j a_j(X(t)) + \sum_{j=1}^M \nu_j \sqrt{a_j(X(t))} W_j(t)$$

In this case, the *discrete stochastic* process $X(t)$ is approximated as a *continuous stochastic* process. But this approximation can be done when all molecular species are sufficiently large.

Furthermore, in the thermodynamic limit [6], the approximation can be taken on a higher level. Since the stochastic model is given by:

$$X(t + \tau) = X(t) + \sum_{j=1}^M \nu_j [N_j(t + \tau) - N_j(t)]$$

where N_j is a counting process and we have $P\{N_j(t + \tau) - N_j(t) \geq 1\} = a_j(X(t))\tau + o(\tau)$. we can make the deterministic approximation:

$$\hat{N}_j(t + \tau) - \hat{N}_j(t) = a_j(X(t))\tau + o(\tau)$$

where \hat{N}_j is now noninteger and deterministic and the corresponding ODE is:

$$\frac{d\hat{X}(t)}{dt} = \sum_{j=1}^M \nu_j a_j(\hat{X}(t))$$

We can see that under certain assumptions, the problem reduces to solving a system of stochastic differential equations(continuous and stochastic regime) and further, in thermodynamic limit(continuous and deterministic approach) the *Explicit Tau Method* leads to the *Explicit Euler method* for the solution of ODE system given by RRE(reaction rate equation).

5 Conclusions and future work

We discussed the explicit tau leaping method for stochastic chemical systems, and we presented a strategy for adaptive step size selection based on local error formulas. Also, relations with the stochastic differential equations and ordinary differential equations were presented, but how to choose the appropriate spectrum is subject of the future work.

References

- [1] Y. CAO, L. R. PETZOLD, M. RATHINAM, AND D.T. GILLESPIE, *The numerical stability of leaping methods for stochastic simulation of chemically reacting systems*, J. Chem. Phys., 121 (2004), pp. 12169–12178.
- [2] D.T. GILLESPIE, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434.
- [3] D.T. GILLESPIE, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., 81 (1977), pp. 2340–2361.
- [4] D. T. GILLESPIE, *Approximate accelerated stochastic simulation of chemically reacting systems*, J. Chem. Phys., 115 (2001), pp. 1716–1733.
- [5] M. B. ELOWITZ, AND S. LEIBLER, *A synthetic oscillatory network of transcriptional regulators*, Nature, 403 (2000), pp. 335–338.
- [6] S. N. ETHIER, AND T. G. KURTZ, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New-York , 1986.
- [7] W. PRESS, B. FLANNERY, S. TEUKOLSKY, AND W. VETTELING, *Numerical Recipes, The Art of Scientific Computing*, 2nd ed., Cambridge University Press, New York , 1986.
- [8] M. RATHINAM, L.R. PETZOLD, Y. CAO, AND D.T. GILLESPIE, *Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method*, J. Chem. Phys., 119 (2003), pp. 11784–1794.
- [9] M. RATHINAM, L.R. PETZOLD, Y. CAO, AND D.T. GILLESPIE, *Consistency and stability of tau leaping schemes for chemical reaction systems*, SIAM J. Multiscale Model. Simul., 4 (2005), pp. 867–895.
- [10] S. ROSS, *Introduction to probability Models*, 8th ed., Academic Press, San Diego, 2002.
- [11] L. F. SHAMPINE, AND M. W. REICHEL, *The Matlab ODE suite*, SIAM J. Sci. Comput., 18 (1997), pp. 1–22.