TRAPEZOIDAL TAU-LEAPING FORMULA FOR THE STOCHASTIC SIMULATION OF BIOCHEMICAL SYSTEMS

Yang Cao and Linda Petzold
Department of Computer Science, University of California Santa Barbara, Santa Barbara, CA 93106.

Abstract
A new formula for accelerated discrete efficient stochastic simulation of chemically reacting system is proposed. This formula has better accuracy and stiff stability properties than the explicit [J. Chem. Phys, 115. 1716-1733, (2001)] and implicit [J. Chem. Phys, 119. 11784-94, (2004)] tau-leaping formulas for discrete stochastic systems, and it limits to the trapezoidal rule in the deterministic regime. Numerical results are presented to demonstrate the advantage of the new formula.

Keywords:
Tau-leaping method, Stiff system, Trapezoidal method, Schlögl reaction.

Introduction

In recent years, concerns over stochastic effects resulting from the small numbers of certain reactant molecules in microscopic systems (Mcadams and Arkin 1997, 1999, Arkin et. al. 1998, Fedoroff and Fontana 2002), have called for accurate and efficient stochastic simulation methods. The fundamental simulation method is Gillespie's Stochastic Simulation Algorithm (SSA) (Gillespie 1976, 1977). Although progress (Gibson 2000, Cao et. al. 2004) has been made to improve the efficiency of implementations of the SSA method, as an exact procedure that simulates every reaction, it is necessarily inefficient for most realistic problems. The tau-leaping method (Gillespie 2001) has been proposed to improve the efficiency. By using a Poisson approximation, the tau-leaping method can *"leap over"* many fast reactions and approximate the stochastic behavior of the system very well. The tau-leaping method makes a natural connection between the SSA in the discrete stochastic regime and the explicit Euler method applied to the chemical Langevin equation in the continuous stochastic regime and to the reaction rate equation (RRE) in the continuous deterministic regime. Thus the tau-leaping method provides a numerical formula ideal for multiscale stochastic simulation.

Several extensions (Gillespie and Petzold 2003, Rathinam et. al. 2003, Cao et. al. 2004) to the tau-leaping method have recently been proposed, including a more robust stepsize selection strategy (Gillespie and Petzold 2003) and an implicit tau-leaping method (Rathinam et. al. 2003). The implicit tau formula is a modification to the tau-leaping method (also called explicit tau formula) to improve the numerical stability of the tau-leaping method when it is applied to stiff (stiffness reflects the presence of multiple timescales, the fastest of which are stable) stochastic systems. The convergence and stability of the explicit and implicit tau methods have been studied (Rathinam et. al. 2004). Despite this progress, research on tau-leaping methods is still at the early stage. Several important problems still remain open. For example: (1) Both the explicit and implicit tau methods are of convergence order one. Higher order formulas are desired. (2) The stepsize selection strategy proposed in Gillespie (2001) and improved in Gillespie and Petzold (2003) is still not as efficient or reliable as it might be. Better stepsize selection strategies are still needed. (3) For stiff problems, the implicit tau method has a damping effect (Rathinam et. al. 2003): the variance generated from the implicit tau formula is much smaller than the exact value for large stepsizes. Although the down-shifting strategy (Rathinam et. al. 2003) can be applied to overcome the damping effect, it would be desirable to have a formula that maintains the stiff stability without the damping effect.

In this paper we propose the trapezoidal tau-leaping formula. This formula is similar to the trapezoidal rule for solving ordinary differential equations (ODEs) (Ascher and Petzold 1998). We will present this formula, and show through numerical experiments that, just as the trapezoidal rule is a second-order formula in the deterministic regime, the trapezoidal tau-leaping formula also is second order for the mean value (although it is still first order for the variance). The numerical experiments also show that, for stiff problems, the trapezoidal tau-leaping formula doesn't have the damping effect. Thus this formula can capture the variance without the use of down-shifting.

Simulation Algorithms for Chemical Kinetics

Suppose the system involves N molecular species $\{S_1, \ldots, S_N\}$. The state vector will be denoted by $X(t) = (X_1(t), \ldots, X_N(t))$, where $X_i(t)$ is the number of molecules of species S_i in the system at time t. M reaction channels $\{R_1, \ldots, R_M\}$ are involved in the system. We assume that the system is well-stirred and in thermal equilibrium. The dynamics of reaction channel R_j is characterized by the *propensity function* a_j and by the *state change vector* $\nu_j = (\nu_{1j}, \ldots, \nu_{Nj})$: $a_j(x)dt$ gives the probability that one R_j reaction will occur in the next infinitesimal time interval $[t, t + dt)$, and ν_{ij} gives the change in the S_i molecular population induced by one R_j reaction.

The dynamics of the system obeys the *chemical master equation* (CME) (Gillespie 1976, 1977). But the CME is hard to solve, both theoretically and numerically. An

equivalent simulation method is the SSA (Gillespie 1976, 1977). Let $a_0(x) = \sum_{j=1}^{M} a_j(x)$. The time τ to the next occurring reaction is the exponentially distributed random variable with mean $1/a_0(x)$. The index j of that reaction is the integer random variable with probability $a_j(x)/a_0(x)$. SSA is a Monte Carlo method based on these distributions. For each step, SSA generates two random numbers r_1 and r_2 in $U(0,1)$, the uniform distribution on the interval $(0,1)$. The time for the next reaction to occur is given by $t+\tau$, where τ is given by

$$\tau = \frac{1}{a_0(t)} \log(\frac{1}{r_1}). \tag{1}$$

The index j for the next reaction is given by the smallest integer satisfying

$$\sum_{l=1}^{j} a_l(t) > r_2 a_0(t). \tag{2}$$

The system states are updated by $X(t+\tau) = X(t) + \nu_j$. Then the simulation proceeds to the next occurring time, until it reaches the final time.

In principle, the SSA could be used to simulate all of the chemical species and reactions, except that because it must proceed one reaction at a time, it is much too slow for most practical problems. Gillespie (2001) has proposed a scheme called *tau-leaping* to accelerate the SSA. The basic idea of the tau-leaping method is to ask the question: How many times does each reaction channel fire in each subinterval? In each step, the tau-leaping method can proceed with many reactions. i This is achieved at the cost of some accuracy. For each j from 1 to M, define $K_j(\tau; x, t)$ as the number of times, given $X(t) = x$, that reaction channel R_j will fire in the time interval $[t, t+\tau)$. Tau-leaping assumes the **Leaping Condition**: *Require τ to be small enough that the change in the state during $[t, t+\tau)$ will be so slight that no propensity function will suffer an appreciable change in its value.* $K_j(\tau; x, t)$ is then given by the Poisson random variable

$$K_j(\tau; x, t) = P(a_j(x)\tau) \quad (j = 1, \ldots, M). \tag{3}$$

The basic tau-leaping method proceeds as follows: Choose a value for τ that satisfies the Leaping Condition. Generate for each $j = 1, \ldots, M$ a sample value k_j of the Poisson random variable $P(a_j(x), \tau)$, and update the state by

$$X(t+\tau) = x + \sum_{j=1}^{M} k_j \nu_j. \tag{4}$$

The tau-leaping method is equivalent to the explicit Euler formula in the deterministic regime. But the explicit Euler formula is known to be inefficient when applied to stiff problems. The basic tau-leaping formula (also called explicit tau-leaping formula) has the same difficulty. The implicit tau formula (Rathinam et. al. 2003) has been proposed to handle the stiffness, and is given by

$$\hat{X}^{(it)}(t+\tau) = x + \sum_{j=1}^{M} \nu_j \left[P(a_j(x)\tau) - a_j(x)\tau + a_j\left(\hat{X}^{(it)}(t+\tau)\right)\right]. \tag{5}$$

Newton's method is used to solve (5) for $X^{(it)}(t+\tau)$. Note that here $X^{(it)}(t+\tau)$ are floating point values. In the simulation, we change them to integers by rounding the quantity in brackets on the right side of (5) to the nearest integer. But to simplify the analysis, here we will use (5) as written. It has been demonstrated (Rathinam et. al. 2003) that the implicit tau formula allows much larger stepsizes than the explicit tau formula, when applied to stiff stochastic systems. The damping effect (Rathinam et. al. 2003) has also been observed for the implicit tau formula: When a large stepsize is used to solve a stiff system, although the implicit tau formula captures the mean value very well, it yields a much smaller variance than the exact value. The variance can be recovered by the "down-shifting" procedure (Rathinam et. al. 2003), but this adds to the time and complexity of the numerical solution procedure.

Trapezoidal Tau-Leaping Formula

It is well-known in the numerical analysis of ODEs that the trapezoidal rule has the following features compared with both the explicit and implicit Euler formulas: (1) The trapezoidal rule has a higher convergence order (two) than the explicit and implicit Euler formulas (which are first order accurate); (2) The trapezoidal rule is A-stable (Ascher and Petzold 1998); (3) The trapezoidal rule does not have the damping effect. We might hope that the analogue of the trapezoidal rule in the discrete stochastic regime might inherent some of these properties.

We propose the trapezoidal tau-leaping formula as follows:

$$\hat{X}^{(tr)}(t+\tau) = x + \sum_{j=1}^{M} \nu_j \left[P(a_j(x)\tau) - \frac{\tau}{2} a_j(x) + \frac{\tau}{2} a_j\left(\hat{X}^{(tr)}(t+\tau)\right)\right]. \tag{6}$$

The difference between (6) and (5) is only the coefficients of the $a_j(x)$ and $a_j(\hat{X})$ terms. The other implementation details are similar to that of the implicit tau formula. After generating the Poisson random numbers $P(a_j(x)\tau)$, we apply Newton's method to solve (6) for $\hat{X}^{(tr)}$. Then K_j is given by

$$K_j^{(tr)} = \text{ROUND}\left(\left[P(a_j(x)\tau) - \frac{\tau}{2} a_j(x) + \frac{\tau}{2} a_j\left(\hat{X}^{(tr)}(t+\tau)\right)\right]\right). \tag{7}$$

The solution of the trapezoidal tau-leaping formula is given by (4), with K_j from (7). For a fixed stepsize τ, the trapezoidal tau formula is formulated as follows:

Algorithm Trapezoidal Tau Method

1. Initialization. (Set the initial numbers of molecules. Set $t = 0$. Set τ.)

2. Calculate the propensity functions $a_i, (i = 1, ..., M)$.

3. Generate M independent random numbers from $P(a_j(x)\tau)$.

4. Solve (6) *for* $\hat{X}^{(tr)}$ *by Newton iteration.*

5. Calculate $K_i, (i = 1, ..., M)$ *by* (7).

6. Update the states of the species by (4). *Update* $t = t + \tau$.

7. End the simulation when t *reaches the final time. Otherwise, go to step 2.*

Numerical Examples

We use three examples to demonstrate the advantages of the trapezoidal tau formula.

Example 1: Isomerization reaction

Consider the isomerization example

$$X \longrightarrow \emptyset, \tag{8}$$

with propensity function $a(x) = 0.1x$, $X(0) = 100$ in the time interval $[0, 10]$. With these parameters, this problem is nonstiff.

For this example we can compute the exact values for the mean and variance and compare them with the numerical results. Figures 1 and 2 show the errors in the mean and variance for different formulas and stepsizes. It is apparent that the errors in the mean from the trapezoidal tau formula are much smaller than the corresponding errors from the explicit and implicit tau formulas. But the errors in the variance from the trapezoidal tau formula show only a linear convergence. By a similar analysis to Rathinam et. al. (2004), we can prove that trapezoidal formula is order two for the mean value and only order one for the variance. This numerical results match with the theoretical analysis.

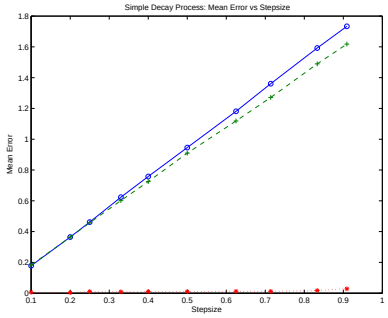

FIG. 1: The errors of the mean vs. the stepsize for Example 1 with 10^6 samples computed using explicit tau (the solid line plot with "o"), the implicit tau (the dashed line with "+"), and the trapezoidal tau (the dotted line with "*").

Example 2: Decaying-dimerizing reaction set

This reaction system was studied in Gillespie (2001) and Rathinam et. al. 2003 as a very stiff system. It consists of $N = 3$ species undergoing $M = 4$ different types of chemical reactions:

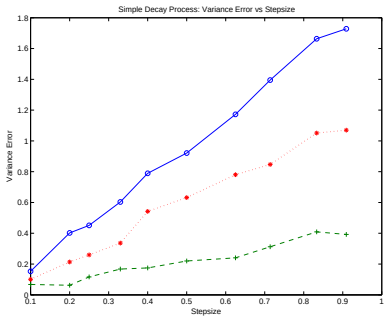

FIG. 2: The errors of the variance vs. the stepsize for Example 1 with 10^6 samples computed using explicit tau (the solid line plot with "o"), the implicit tau (the dashed line with "+"), and the trapezoidal tau (the dotted line with "*").

$$\begin{aligned} S_1 &\xrightarrow{c_1} 0 \\ S_1 + S_1 &\underset{c_3}{\overset{c_2}{\rightleftharpoons}} S_2 \\ S_2 &\xrightarrow{c_4} S_3. \end{aligned} \tag{9}$$

We chose values for the parameters

$$c_1 = 1,\ c_2 = 10,\ c_3 = 1000,\ c_4 = 0.1.$$

Hence the propensity functions are given by

$$a_1 = x_1,\ a_2 = 5x_1(x_1 - 1),\ a_3 = 1000x_2,\ a_4 = 0.1x_2,$$

x_i being the number of species S_i. We chose the initial conditions $X_1(0) = 400$, $X_2(0) = 798$, $X_3(0) = 0$ and final time $T = 0.2$. Table I shows the mean and variance values of $10,000$ samples generated independently by different formulas with different stepsizes compared with the results of $10,000$ samples from SSA. As we can see, the explicit tau formula blows up for larger stepsizes. The implicit tau formula produces an accurate mean but too small variance, due to the damping effect. The trapezoidal tau formula generates an accurate variance even with large stepsizes.

Method	*Stepsize*	10^{-5}	10^{-4}	2×10^{-4}	10^{-3}	10^{-2}
Explicit	*Mean*	*387.68*	*386.60*	*384.85*	∞	∞
Tau	*Variance*	*362.47*	*598.48*	2486	∞	∞
Implicit	*Mean*	*387.18*	*387.33*	*387.57*	*387.60*	*387.55*
Tau	*Variance*	*345.90*	*242.17*	*187.87*	*64.56*	*9.25*
Trapezoidal	*Mean*		*387.30*	*387.33*	*387.43*	*387.79*
Tau	*Variance*		*347.49*	*352.50*	*347.64*	*338.92*

TABLE I: The mean and variance for X_1 at $T = 0.2$, generated from different formulas. The SSA values are: mean 387.6931 and variance 343.5377.

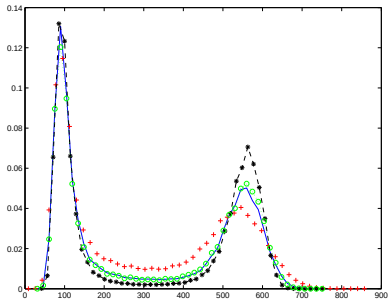

FIG. 3: The histogram (10,000 samples) of X solved by the SSA method (solid line), explicit tau (plot with '+'), implicit tau (plot with '*') and trapezoidal method (plot with 'o') for Schlögl reaction.

Example 3: Schlögl reaction: a bistable example

The Schlögl reaction (Gillespie 1992) is a famous example of a reactor with bistable distribution.

$$\begin{aligned} B_1 + 2X &\underset{c_2}{\overset{c_1}{\rightleftharpoons}} 3X, \\ B_2 &\underset{c_4}{\overset{c_3}{\rightleftharpoons}} X. \end{aligned} \tag{10}$$

B_1 and B_2 denote buffered species whose molecular populations N_1 and N_2 are assumed to remain essentially constant over the time interval of interest. Let

$$X(t) = \text{number of } X \text{molecules in the system at time } t.$$

The propensity functions are

$$\begin{aligned} a_1(x) &= c_1 N_1 x(x-1), \\ a_2(x) &= c_2 x(x-1)(x-2), \\ a_3(x) &= c_3 N_2, \\ a_4(x) &= c_4 x. \end{aligned} \tag{11}$$

The state change vectors are $\nu_1 = \nu_3 = 1$, $\nu_2 = \nu_4 = -1$. For some parameter values, this reaction has two stable states. The special parameter set that we used in our simulation is:

$$\begin{gathered} c_1 = 3 \times 10^{-7}, \quad c_2 = 10^{-4}, \quad c_3 = 10^{-3}, \quad c_4 = 3.5; \\ N_1 = 1 \times 10^5, \quad N_2 = 2 \times 10^5. \end{gathered} \tag{12}$$

We ran the simulation from $t = 0$ with initial state $x(0) = 250$ to time $T = 4$. The histograms generated from SSA, explicit tau, implicit tau and trapezoidal tau with fixed stepsize $\tau = 0.4$ are shown in Figure 3. We can see that the histogram given by the trapezoidal tau method is much closer to that given by the SSA method than those given by the explicit and implicit tau methods.

Conclusions

A new tau-leaping formula for discrete stochastic simulation of biochemical systems was introduced. Numerical experiments verify it excellent stability and accuracy properties.

Acknowledgments

This work was supported by the California Institute of Technology under DARPA Award No. F30602-01-2-0558, by the U. S. Department of Energy under DOE award No. DE-FG02-04ER25621, by the National Science Foundation under NSF award CCF-0326576 and ACI00-86061, and by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U. S. Army Research Office.

References

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

Ascher U. and Petzold L. (1998). *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations.* SIAM.

Cao Y., Li H. and Petzold L. (2004). Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chem. Phys.*, 121, 4059–67.

Cao Y., Petzold L., Rathinam L. and Gillespie D. (2004). The numerical stability of leaping methods for stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 121, 12169–78.

Fedoroff N. and Fontana W. (2002). Small numbers of big molecules. *Science*, 297, 1129–1131.

Gibson M. and Bruck J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104, 1876.

Gillespie D. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22, 403–434.

Gillespie D. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81, 2340–61.

Gillespie D. (1992). *Markov Processes: An Introduction for Physical Scientists.* Academic Press.

Gillespie D. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115, 1716.

Gillespie D. and Petzold L. (2003). Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.*, 119.

McAdams H. and Arkin A. (1997). Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*, 94, 814–819.

McAdams H. and Arkin A. (1999). It's a noisy business! *Trends in Genetics*, 15(2), 65–69.

Rathinam M, Petzold L., Cao Y., Gillespie D. (2003). Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. *J. Chem. Phys.*, 119, 12784–94.

Rathinam M, Petzold L. , Cao Y., Gillespie D. (2005). Consistency and stability of tau leaping schemes for chemical reaction systems. *to appear, SIAM Multiscale.*