Binomial leap methods for simulating stochastic chemical kinetics

Tianhai Tian and Kevin Burrage

Citation: The Journal of Chemical Physics **121**, 10356 (2004); doi: 10.1063/1.1810475
View online: http://dx.doi.org/10.1063/1.1810475
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/121/21?ver=pdfcov
Published by the AIP Publishing

Binomial leap methods for simulating stochastic chemical kinetics

Tianhai Tian[a] and Kevin Burrage[b]
Advanced Computational Modelling Centre, University of Queensland, Brisbane QLD 4072, Australia

(Received 29 April 2004; accepted 3 September 2004)

This paper discusses efficient simulation methods for stochastic chemical kinetics. Based on the τ-leap and midpoint τ-leap methods of Gillespie [D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001)], binomial random variables are used in these leap methods rather than Poisson random variables. The motivation for this approach is to improve the efficiency of the Poisson leap methods by using larger stepsizes. Unlike Poisson random variables whose range of sample values is from zero to infinity, binomial random variables have a finite range of sample values. This probabilistic property has been used to restrict possible reaction numbers and to avoid negative molecular numbers in stochastic simulations when larger stepsize is used. In this approach a binomial random variable is defined for a single reaction channel in order to keep the reaction number of this channel below the numbers of molecules that undergo this reaction channel. A sampling technique is also designed for the total reaction number of a reactant species that undergoes two or more reaction channels. Samples for the total reaction number are not greater than the molecular number of this species. In addition, probability properties of the binomial random variables provide stepsize conditions for restricting reaction numbers in a chosen time interval. These stepsize conditions are important properties of robust leap control strategies. Numerical results indicate that the proposed binomial leap methods can be applied to a wide range of chemical reaction systems with very good accuracy and significant improvement on efficiency over existing approaches. © *2004 American Institute of Physics.*
[DOI: 10.1063/1.1810475]

I. INTRODUCTION

Stochastic modeling of biological systems has become a very important research field in recent years. Experimental and theoretical studies have shown the importance of stochastic processes in genetic regulatory networks and cellular processes.[1–5] For biological systems involving molecules of small populations, the stochastic simulation algorithm (SSA) derived by Gillespie[6] is an essentially exact procedure for studying noise in chemical kinetic systems. However, the computational load of the SSA is often very high when it is applied to simulate large biological systems. Thus it is imperative to design efficient numerical methods for simulating stochastic chemical kinetics.

There are two significant approaches for reducing the computational time of the SSA. The first approach is based on a new approach of Gillespie through the use of leap methods with Poisson random variables.[7] In the Poisson τ-leap method a number of reactions are allowed to fire in a relative larger time interval rather than a single reaction firing in the next-reaction time interval, as is the case of the SSA. Following the Poisson τ-leap method, the midpoint τ-leap method,[7] implicit τ-leap method,[8] and Poisson Runge–Kutta methods[9] have been designed recently in order to improve the accuracy and efficiency of the simulations. However, robust leap control strategies should be developed before these methods can be considered for practical applications.[7] Recently Gillespie and Petzold have presented an improved leap size selection procedure for determining the maximum leap size for a specified degree of accuracy.[10]

The second approach is to partition a chemical reaction system into subsets of slow and fast reactions and then to apply different simulation methods to each subset. Rao and Arkin demonstrated how to reduce computational time by applying the quasisteady state assumption to the subset of fast reactions.[11] Haseltine and Rawlings improved the computational efficiency by approximating fast reactions either deterministically or as Langevin equations.[12] The open problem in the second approach is how to simulate chemical reactions with reactant species of intermediate molecular numbers and/or with intermediate values of propensity functions. Recently Burrage *et al.*[13] partitioned chemical reaction systems into three subsets of slow, intermediate, and fast reactions and used the Poisson τ-leap method to simulate the subset of intermediate reactions. The improvement over the SSA implementation is substantial rather than dramatic. The complexity of the partitioning process eroded potential efficiency gains. In addition to the methods mentioned above, other methods have also been proposed recently, for example, Gibson and Bruck's method with less required random numbers,[14] Gillespie's continuous model[15] and the probability-weighted Monte Carlo approach by Resat *et al.*[16]

In this paper we will use binomial random variables in the leap methods instead of Poisson random variables. This is not intended just to provide an alternative sampling from a Poisson distribution but will also address the issues of robust leap control strategies. It will be seen that the proposed binomial leap methods are robust and very efficient for simulating chemical reaction systems. The rest of this paper is

[a]Electronic mail: tian@maths.uq.edu.au
[b]Electronic mail: kb@maths.uq.edu.au

organized as follows. We give a brief review of the SSA and Poisson leap methods in Sec. II, and present the Binomial leap methods in Sec. III. In Secs. IV–VI we report the accuracy and efficiency of binomial leap methods by simulating three reaction systems, namely, the isomerization reaction in Sec. IV, a small system with four reaction channels in Sec. V and the expression and activity of LacY and LacZ in *E. coli* in Sec. VI.

II. STOCHASTIC SIMULATION ALGORITHMS

In this paper we will study the evolution of molecular numbers in a well stirred chemical reaction system. This system contains N molecular species $\{S_1,\ldots,S_N\}$ with number $X_i(t)$ of the species S_i at time t. These species of molecules chemically interact inside some fixed volume Ω at a constant temperature through reaction channels $\{R_1,\ldots,R_M\}$.

For each reaction channel R_j $(j=1,...,M)$, we define a propensity function $a_j(\mathbf{x})$ in a given state $\mathbf{X}(t)=[X_1(t),\ldots,X_N(t)]^{\mathrm{T}}=\mathbf{x}$ and use $a_j(\mathbf{x})dt$ to represent the probability that one reaction R_j will occur somewhere inside Ω in the infinitesimal time interval $[t,t+dt)$. In addition a state change vector ν_j is defined to characterize reaction channel R_j. The element ν_{ij} of ν_j represents the change in the number of species S_i due to reaction R_j.

The SSA is any statistically exact procedure for generating the time and index of the next occurring reaction in accordance with the current values of the propensity functions. In the so-called direct method,[7] we draw two independent random numbers r_1 and r_2 from the uniform distribution in the unit interval, and then take the time of the next reaction to be the current time plus μ, where

$$\mu=\frac{1}{a_0(\mathbf{x})}\ln\left(\frac{1}{r_1}\right)$$

and the index of the next reaction to be the value of j that satisfies

$$\sum_{k=1}^{j-1} a_k(\mathbf{x})<r_2 a_0(\mathbf{x})\leqslant\sum_{k=1}^{j} a_k(\mathbf{x}).$$

Here $a_0(\mathbf{x})=\Sigma_{k=1}^{M}a_k(\mathbf{x})$. Then the system is updated by

$$\mathbf{x}(t+\mu)=\mathbf{x}(t)+\nu_j.$$

It is assumed in the Poisson τ-leap method that there are a number of reactions firing in a relatively larger time interval $[t,t+\tau)$. The reaction number of channel R_j firing in $[t,t+\tau)$ is a sample value generated from a Poisson random variable $\mathcal{P}[a_j(\mathbf{x})\tau]$ with mean $a_j(\mathbf{x})\tau$. The probability function of $\mathcal{P}[a_j(\mathbf{x})\tau]$ is

$$\Pr\{\mathcal{P}[a_j(\mathbf{x})\tau]=K\}=\frac{[a_j(\mathbf{x})\tau]^K}{K!}e^{-a_j(\mathbf{x})\tau},\quad K=0,1,...,\infty.$$

After generating a sample values K_j from $\mathcal{P}[a_j(\mathbf{x})\tau]$ for each reaction channel, the system is updated by

$$\mathbf{x}(t+\tau)=\mathbf{x}(t)+\sum_{j=1}^{M}\nu_j K_j. \tag{1}$$

The stepsize τ should satisfy the leap condition[10]

$$\tau=\min_{j\in[1,M]}\left\{\frac{\epsilon a_0(\mathbf{x})}{|\mu_j(\mathbf{x})|},\frac{\epsilon^2 a_0^2(\mathbf{x})}{\sigma_j^2(\mathbf{x})}\right\}, \tag{2}$$

where

$$f_{jk}(\mathbf{x})=\sum_{i=1}^{N}\frac{\partial a_j(\mathbf{x})}{\partial x_i}\nu_{ik},\quad j,k=1,...,M,$$

$$\mu_j(\mathbf{x})=\sum_{k=1}^{M}f_{jk}(\mathbf{x})a_k(\mathbf{x}),\quad j=1,...,M,$$

$$\sigma_j^2(\mathbf{x})=\sum_{k=1}^{M}f_{jk}^2(\mathbf{x})a_k(\mathbf{x}),\quad j=1,...,M.$$

This procedure attempts to ensure that the change in each propensity function during a leap of size τ will be no larger than $\epsilon a_0(\mathbf{x})$, where ϵ is a prespecified error control parameter $(0<\epsilon\ll 1)$. In addition, it would be better to forego the leap strategy and instead use the SSA if the determined stepsize τ is less than a few multiples of $1/a_0(\mathbf{x})$. The SSA will be used if the selected leap size satisfies

$$\tau\leqslant\frac{k}{a_0(\mathbf{x})}, \tag{3}$$

where k can be any number between 1 and 10.[7]

In the Poisson τ-leap method, state $\mathbf{x}(t)$ is used to approximate the states of the system in the time interval $[t,t+\tau)$. In order to improve the accuracy, a predicted state at a point in $[t,t+\tau)$ can be used to approximate the states of the system. Similar to the midpoint Runge–Kutta method for solving ordinary differential equations, a predicted state at the midpoint $(t+\tau/2)$ is defined by

$$\mathbf{x}=\mathbf{x}+\left\lfloor\frac{1}{2}\tau\sum_{j=1}^{M}a_j(\mathbf{x})\nu_j\right\rfloor, \tag{4}$$

where $\lfloor x\rfloor$ is the largest integer in x. In the Poisson midpoint τ-leap method,[7] a sample value K_j is generated from the Poisson random variable $\mathcal{P}[a_j(\bar{\mathbf{x}})\tau]$ for each $j=1,...,M$ and the system is updated by

$$\mathbf{x}(t+\tau)=\mathbf{x}(t)+\sum_{j=1}^{M}\nu_j K_j. \tag{5}$$

The Poisson τ-leap and midpoint τ-leap methods are special cases of the following s-stage Poisson Runge–Kutta methods,[9] defined by

$$\begin{aligned}Y_i&=\mathbf{x}(t)+\sum_{k=1}^{M}\nu_k\mathcal{P}\left[\sum_{j=1}^{s}w_{ij}a_k(Y_j)\tau\right],\quad i=1,\cdots,s,\\ \mathbf{x}(t+\tau)&=\mathbf{x}(t)+\sum_{k=1}^{M}\nu_k\mathcal{P}\left[\sum_{j=1}^{s}\beta_j a_k(Y_j)\tau\right].\end{aligned} \tag{6}$$

In addition, the Heun and $R2$ methods have also been presented by Burrage and Tian.[9]

III. BINOMIAL LEAP METHODS

When applying the Poisson τ-leap methods to stochastic chemical kinetics, we should be very careful about stepsize

selection. Negative molecular numbers may be obtained if certain species have small molecular numbers and the stepsize is large. Recently, Gillespie and Petzold have proposed a procedure (2) for improving the robustness of stochastic simulations.[10] Instead of using more cautious stepsize selection procedures, we will propose another approach in order to improve the efficiency of the Poisson τ-leap method by using larger stepsizes. Probabilistic properties of random variables will be used to restrict the reaction numbers and to avoid possible negative molecular numbers when a large stepsize is used.

There are two ways of obtaining negative molecular numbers in stochastic simulations. The first way is that the sample value for the reaction number may be greater than one of the molecular numbers in that reaction channel. For example, consider a reaction

$$S_1+S_2\overset{c_1}{\rightarrow}S_3 \tag{7}$$

with $x_1=1000$, $x_2=1$, and $c_1=0.1$. The reaction number in the time interval $[t,t+\tau)$ is a sample value of the Poisson random variable $\mathcal{P}(c_1x_1x_2\tau)=\mathcal{P}(100\tau)$ in the Poisson τ-leap method. If reaction (7) is one of the reactions in a system with large $a_0(\mathbf{x})$, it is possible to generate a sample value that is greater than 1. An example of this possibility can be found in Sec. VI.

The second case can arise due to the simultaneous occurrence of different reaction channels. For example, if a system contains reaction channels

$$S_1+S_2\overset{c_1}{\rightarrow}S_3,$$

$$S_1\overset{c_2}{\rightarrow}S_4,$$

the total reaction number of these two reaction channels may be greater than the molecular number of species S_1 even if the reaction number of each channel is less than the number of S_1.

For tackling the problem of negative numbers, we introduce binomial random variables to restrict the possible reaction numbers in the next time interval. A binomial random variable $\mathcal{B}(N,p)$ denotes N repeated independent Bernoulli trials and each trial has probability of success p. A sample value of $\mathcal{B}(N,p)$ is a integer between 0 and N. This finite range of sample values allows us to properly bound the numbers of reactions and avoid negative populations. In addition, Poisson and binomial random numbers are simular to each other. The probability function of $\mathcal{B}(N,p)$ is

$$\Pr[\mathcal{B}(N,p)=K]=\frac{N!}{K!(N-K)!}p^K(1-p)^{N-K},$$

$$K=0,1,...,N.$$

The mean of $\mathcal{B}(N,p)$ is Np that equals to the mean of $\mathcal{P}(Np)$. If N is large and p is small, a binomial random variable $\mathcal{B}(N,p)$ can be approximated by a Poisson random variable $\mathcal{P}(Np)$.

In the Poisson τ-leap method, the reaction number of channel R_j in the time interval $[t,t+\tau)$ is a sample value of the Poisson random variable $\mathcal{P}[a_j(\mathbf{x})\tau]$. In the binomial τ-leap method introduced in this paper, the reaction number of channel R_j is defined by a sample value of the binomial random variable $\mathcal{B}[N_j,a_j(\mathbf{x})\tau/N_j]$ under the condition

$$0\leqslant\frac{a_j(\mathbf{x})\tau}{N_j}\leqslant 1. \tag{8}$$

In order to keep positive molecular numbers in stochastic simulations, we define functions N_j below for the widely used three types of elementary reactions.

(1) The first-order reaction

$$S_1\overset{c_1}{\rightarrow}S_3,\quad a_j(\mathbf{x})=c_1x_1,\quad N_j=x_1. \tag{9}$$

(2) The second-order reaction

$$S_1+S_2\overset{c_2}{\rightarrow}S_4,\quad a_j(\mathbf{x})=c_2x_1x_2,\quad N_j=\min\{x_1,x_2\}. \tag{10}$$

(3) The homodimer formation ($x_1\geqslant 2$)

$$S_1+S_1\overset{c_3}{\rightarrow}S_5,\quad a_j(\mathbf{x})=\tfrac{1}{2}c_3x_1(x_1-1),\quad N_j=\lfloor\tfrac{1}{2}x_1\rfloor. \tag{11}$$

According to Eq. (10), the reaction number of the second-order reaction is less than or equal to the smaller of the two molecular numbers. In the case of the homodimer reaction, we use $N_j=\lfloor 1/2x_1\rfloor$ since two molecules are needed for one reaction. For one single reaction, the defined N_j above can ensure positive molecular numbers after one time step.

Now we return to the example reaction (7). Using definition (10), the reaction number now is a sample value of the binomial random variable $\mathcal{B}(x_2,c_1x_1\tau)=\mathcal{B}(1,100\tau)$ under the condition $100\tau\leqslant 1$, which is either 1 or 0.

Next we consider the second issue of obtaining negative molecular numbers. A sampling technique will be designed for the total reaction number of a reactant species that undergoes two or more reaction channels. This technique is based on the following two properties of the Poisson and binomial random variables.

Property 1. If $\mathcal{P}_1=\mathcal{P}(\lambda_1)$ and $\mathcal{P}_2=\mathcal{P}(\lambda_2)$ are two independent Poisson random variables with means λ_1 and λ_2, respectively, then $\mathcal{P}_1+\mathcal{P}_2$ is also a Poisson $\mathcal{P}(\lambda_1+\lambda_2)$ with mean $\lambda_1+\lambda_2$.

Property 2. If $\mathcal{P}_1=\mathcal{P}(\lambda_1)$ and $\mathcal{P}_2=\mathcal{P}(\lambda_2)$ are two independent Poisson random variables with means λ_1 and λ_2, respectively, then the conditional probability $\Pr(\mathcal{P}_1=K_1|\mathcal{P}_1+\mathcal{P}_2=K)$ equals to the probability $\Pr(\mathcal{B}=K_1)$ of a Binomial random variable $\mathcal{B}=\mathcal{B}(K,\lambda_1/(\lambda_1+\lambda_2))$ for $K_1=0,1,...,K$.

Property 1 can be found in a textbook of probability theory. For $K_1=0,1,...,K$, property 2 can be derived as follows:

$$\Pr(\mathcal{P}_1=K_1|\mathcal{P}_1+\mathcal{P}_2=K)$$
$$=\frac{\Pr(\mathcal{P}_1=K_1)\Pr(\mathcal{P}_2=K-K_1)}{\Pr(\mathcal{P}_1+\mathcal{P}_2=K)}$$
$$=\left[\frac{e^{-\lambda_1}}{K_1!}\lambda_1^{K_1}\frac{e^{-\lambda_2}}{(K-K_1)!}\lambda_2^{K-K_1}\right]\Bigg/\left[\frac{e^{-\lambda_1-\lambda_2}}{K!}(\lambda_1+\lambda_2)^K\right]$$
$$=\frac{K!}{K_1!(K-K_1)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{K_1}\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{K-K_1}$$
$$=\Pr\left[\mathcal{B}\left(K,\frac{\lambda_1}{\lambda_1+\lambda_2}\right)=K_1\right].$$

Now we consider a sampling technique for generating reaction numbers of two reaction channels R_j and R_k which species S_i undergoes. Similar consideration can be given to three or more simultaneous reaction channels. Let the propensity functions of R_j and R_k be written as

$$a_j(\mathbf{x})=N_j\frac{a_j(\mathbf{x})}{N_j},\quad a_k(\mathbf{x})=N_k\frac{a_k(\mathbf{x})}{N_k},$$

where N_j and N_k, defined by Eqs. (9), (10), or (11), are functions of the population x_i of species S_i. Note that N_j is either given by $N_j=x_i$ or $N_j=\lfloor x_i/2\rfloor$. Here we use the Poisson τ-leap method as the starting point. In the Poisson τ-leap method, reaction numbers of R_j and R_k in $[t,t+\tau)$ are sample values from the Poisson random variables $\mathcal{P}_j=\mathcal{P}[a_j(\mathbf{x})\tau]$ and $\mathcal{P}_k=\mathcal{P}[a_k(\mathbf{x})\tau]$, respectively. According to property 1, the total reaction number of channels R_j and R_k is a sample value K_{jk} from the Poisson random variable

$$\mathcal{P}_j+\mathcal{P}_k=\mathcal{P}\{[a_j(\mathbf{x})+a_k(\mathbf{x})]\tau\}. \tag{12}$$

Then by using property 2, the probability of $\mathcal{P}_j=K_j$, given that the sample value of $\mathcal{P}_j+\mathcal{P}_k$ is K_{jk}, is given by

$$\Pr(\mathcal{P}_j=K_j|\mathcal{P}_j+\mathcal{P}_k=K_{jk})=\Pr\left[\mathcal{B}\left(K_{jk},\frac{a_j(\mathbf{x})}{a_j(\mathbf{x})+a_k(\mathbf{x})}\right)=K_j\right]. \tag{13}$$

In order to keep positive molecular numbers, the total reaction number of R_j and R_k is not generated from the Poission random variable (12) but from a binomial random variable. Similar to the consideration for a single reaction channel, the Poission random variable (12) is replaced by the following binomial random variable:

$$\mathcal{B}\left(N_i,\frac{a_j(\mathbf{x})+a_k(\mathbf{x})}{N_i}\tau\right) \tag{14}$$

under the condition $N_i=\min\{N_j,N_k\}\neq 0$ and

$$0\leqslant\frac{a_j(\mathbf{x})+a_k(\mathbf{x})}{N_i}\tau\leqslant 1. \tag{15}$$

Based on the discussion above we have the following sampling technique for reaction numbers of channels R_j and R_k.

(1) Generate a sample value K_{jk} for the total reaction number of R_j and R_k from the binomial random variable (14);

(2) generate a sample value K_j for the reaction number of R_j from

$$\mathcal{B}\left[K_{jk},\frac{a_j(\mathbf{x})}{a_j(\mathbf{x})+a_k(\mathbf{x})}\right];$$

(3) and the reaction number of channel R_k is $K_k=K_{jk}-K_j$.

Then we can define the Binomial τ-leap method which is given below.

Method 1. *For a given error control parameter ϵ, choose a stepsize τ from the τ-selection process (2), that satisfies stepsize conditions (8) for each reaction channel. Then generate a sample value K_j from the binomial random variable $\mathcal{B}[N_j,a_j(\mathbf{x})\tau/N_j]$ for $j=1,...,M$. If there are reactant species undergoing two or more reaction channels, apply the simultaneous reaction stepsize condition (15) and sampling technique for these reaction channels. Finally update the system by*

$$\mathbf{x}(t+\tau)=\mathbf{x}(t)+\sum_{j=1}^{M}\nu_jK_j.$$

Similarly we can consider the binomial midpoint τ-leap method. It should be noticed that the predicted state $\bar{\mathbf{x}}$ in the Poisson midpoint τ-leap method is used to provide more accurate propensity functions, and the update is based on sample values from the Poisson random variables $\mathcal{P}(a_j(\bar{\mathbf{x}})\tau)$. When using binomial random variables, we have the following two schemes after the midpoint prediction (4).

Scheme 1 Use the predicted state $\bar{\mathbf{x}}$ to define $\bar{N}_j$, and then generate a sample value from the binomial random variable $\mathcal{B}[\bar{N}_j,a_j(\bar{\mathbf{x}})\tau/\bar{N}_j]$.

Scheme 2 Use the state $\mathbf{x}$ at t to define N_j and calculate the propensity function $a_j(\bar{\mathbf{x}})$. Then generate a sample value from the binomial random variable $\mathcal{B}[N_j,a_j(\bar{\mathbf{x}})\tau/N_j]$ under the midpoint prediction conditions

$$N_j\neq 0,\quad 0\leqslant\frac{a_j(\bar{\mathbf{x}})}{N_j}\tau\leqslant 1. \tag{16}$$

In fact $N_j\neq 0$ is a condition for both schemes above. If $N_j=0$, it is difficult to make the midpoint prediction. By using scheme 1, any movement from the number zero to a nonzero $\bar{N}_j$ will lead to either possibly unreasonable sample values ($N_j>0$) or a meaningless binomial random variable ($\bar{N}_j<0$). If $N_j\neq 0$, numerical simulations in Sec. V suggest that it would be better to use scheme 2 although additional time is needed for checking the midpoint prediction conditions (16). Then we have the following binomial midpoint τ-leap method.

Method 2. *Select the leaping time τ by using the τ-selection process (2) and stepsize conditions (8) with a given error control parameter ϵ. Then compute the expected state $\bar{\mathbf{x}}$ (4) at $t+\tau/2$, use the state $\mathbf{x}$ at t to define N_j ($N_j\neq 0$), and generate a sample value K_j of the binomial random variable $\mathcal{B}[N_j,a_j(\bar{\mathbf{x}})\tau/N_j]$ for $j=1,...,M$. If there are reactant species undergoing two or more reaction channels,*

apply the simultaneous reaction stepsize condition (15) and sampling technique for these reaction channels. Finally update the system by

$$\mathbf{x}(t+\tau)=\mathbf{x}(t)+\sum_{j=1}^{M}\nu_j K_j.$$

Random number generation is an important issue for the efficiency of stochastic simulations. Computer programs for generating Poisson and binomial random numbers can be found in Press *et al.*[17] and Netlib (www.netlib.org/random/random.f90) in FORTRAN. The latter is a random number generation library of FORTRAN routines including generators for 14 random variables such as the normal, gamma, Poisson, and binomial random variables.[18,19] Here we recommend to use the generator *random_Poisson* in Press *et al.*[17] for Poisson samples and the function *random_binomial* for binomial samples that is based on the algorithm BTPE.[18]

Functions *poissrnd* and *binornd* in Matlab are also available for generating sample values of the Poisson and binomial random variables, respectively. Computing time (flops in Matlab) for generating Poisson samples is a linear function of the sample value, but for binomial samples, it is a function of the number of trials *N*. Computing time in Matlab normally is large but these two functions can be used to simulate small systems for measuring the accuracy of different methods.

IV. SYSTEM 1: THE ISOMERIZATION REACTION

In this section the isomerization reaction

$$X\xrightarrow{c}Y \tag{17}$$

is used to test the accuracy of different simulation methods. The propensity function of this reaction is $a(x)=cx$ and the state change vector is $\nu=-1$. The solution to the chemical master reaction equation (CMR) is[7]

$$\Pr(x-k,t+\tau|x,t)=\frac{x!}{k!(x-k)!}[1-e^{-c\tau}]^k[e^{-c\tau}]^{x-k},$$
$$(0\leqslant k\leqslant x;\tau\geqslant 0), \tag{18}$$

which is the probability that, given the population x at t, there are k isomerization reactions in the time interval $[t,t+\tau)$.

By using the binomial τ-leap method, the number of X at $t+\tau$ is $x(t+\tau)=x(t)-k$. Here k is a sample value of the binomial random variable $\mathcal{B}(x,c\tau)$ under the stepsize condition $0\leqslant c\tau\leqslant 1$. The density function of $\mathcal{B}(x,c\tau)$ is

$$P_{\mathcal{B}}(k;x,c\tau)=\frac{x!}{k!(x-k)!}(c\tau)^k(1-c\tau)^{x-k},\quad k=0,...,x, \tag{19}$$

which can be regarded as a linear approximation to solution (18) since

$$e^{-c\tau}=1-c\tau+O(c^2\tau^2).$$

For the binomial midpoint τ-leap method, the molecular number at $t+\tau$ is also $x(t+\tau)=x(t)-k$ but the sample value k is from the binomial random variable $\mathcal{B}(x,c\tau\bar{x}/x)$ $(x\neq 0)$ with a predicted state $\bar{x}=x-\lfloor cx\tau x/2\rfloor$. The stepsize condition is $0\leqslant c\tau\bar{x}/x\leqslant 1$ and the density function of $\mathcal{B}(x,c\tau\bar{x}/x)$ is

$$P_{\mathcal{B}}\left(k;x,\frac{c\tau\bar{x}}{x}\right)=\frac{x!}{k!(x-k)!}\left(c\tau-\frac{c\tau}{x}\left\lfloor\frac{cx\tau}{2}\right\rfloor\right)^k\times\left(1-c\tau+\frac{c\tau}{x}\left\lfloor\frac{cx\tau}{2}\right\rfloor\right)^{x-k}, \tag{20}$$

for $k=0,...,x$. If $\lfloor c\tau x/2\rfloor=c\tau x/2$, the density function (20) can be regarded as a second-order approximation to solution (18), namely,

$$e^{-c\tau}=1-c\tau+\tfrac{1}{2}c^2\tau^2+O(c^3\tau^3).$$

When $\lfloor c\tau x/2\rfloor\neq c\tau x/2$, function (20) can still give very good approximation to solution (18) because the difference between $c\tau/x\lfloor c\tau x/2\rfloor$ and $c^2\gamma^2/2$ is small.

Compared with the density function of the Poisson τ-leap method[7]

$$P_{\mathcal{P}}(k;cx\tau)=\frac{(xc\tau)^k}{k!}e^{-xc\tau},\quad k=0,...,\infty \tag{21}$$

and that of the Poisson midpoint τ-leap method[7]

$$P_{\mathcal{P}}(k;c\bar{x}\tau)=\frac{(\bar{x}c\tau)^k}{k!}e^{-\bar{x}c\tau},\quad k=0,...,\infty \tag{22}$$

with $\bar{x}=x-\lfloor c\tau x/2\rfloor$, the density functions (19) and (20) of the binomial leap methods give better approximations to the solution of the CMR (18).

Figure 1 gives probability density functions (18), (19), (20), (21), and (22) for $c=1$, $x=100$, and different stepsizes

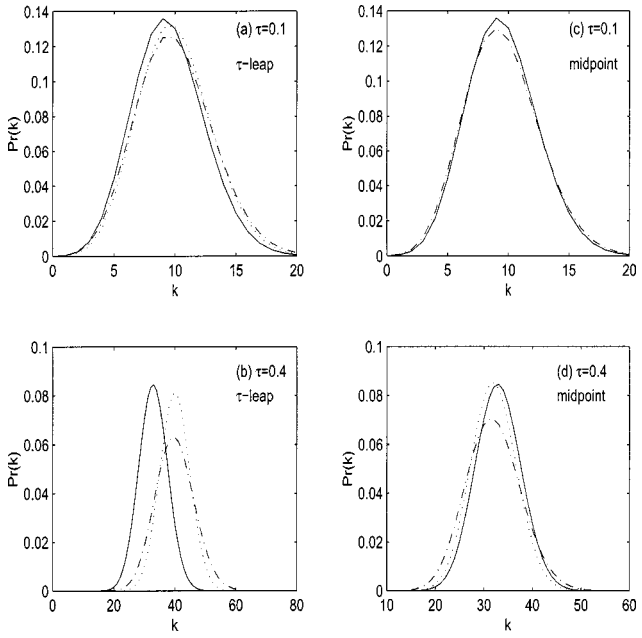

FIG. 1. Probability density functions for the number k of isomerizations in system 1 occurring in a given time τ for $c=1$ and $x=100$. [(a) and (b)] Distributions of the τ-leap methods with Poisson and binomial random variables, respectively. [(c) and (d)] Distributions of the midpoint τ-leap methods with Poisson and binomial random variables, respectively. (Solid line: solution of CMR; dash-line: a Poisson leap method; dot-line: a binomial leap method).

τ. When $\tau=0.1$, both the Poisson and binomial midpoint τ-leap methods give very good approximations to the solution of the CMR [Fig. 1(c)] while the Poisson and binomial τ-leap methods only give acceptable results [Fig. 1(a)]. In addition, there is not any significant difference between the distributions of the binomial leap methods and the corresponding Poisson leap methods. However, for a larger stepsize $\tau=0.4$ [Figs. 1(b) and (d)], the binomial leap methods give better approximations to the variance of the distributions than the corresponding Poisson leap methods, although the Poisson and binomial leap methods have the same shift in the first moment of distributions. For both stepsizes the binomial midpoint τ-leap method gives a better approximation than the binomial τ-leap method. Similar phenomena can be observed for the Poisson leap methods.

Note that the stepsize in any practical computation is quite small. Stepsize τ for simulating reaction (17) with $c=1$ and $x=100$ is $\min\{\epsilon,100\epsilon^2\}$ if the selection process (2) is applied. For the systems in Secs. V and VI, values of $\epsilon=0.05$, 0.03, or 0.01 are used in order to attain good approximations.

V. SYSTEM 2: A SYSTEM WITH FOUR REACTION CHANNELS

The second test system contains three reactant species and four reaction channels, defined by

$$\begin{aligned} R_1&: \quad S_1 \xrightarrow{c_1} (),\\ R_2&: \quad S_1+S_1 \xrightarrow{c_2} S_2,\\ R_3&: \quad S_2 \xrightarrow{c_3} S_1+S_1,\\ R_4&: \quad S_2 \xrightarrow{c_4} S_3. \end{aligned} \tag{23}$$

Detailed simulations of this system in the time interval [0,30] can be found in Gillespie[7] and Burrage and Tian[9] based on the initial condition $\mathbf{x}(0)=(10^5,0,0)^{\mathrm{T}}$ and rate constants $\mathbf{c}=(1,0.002,0.5,0.04)^{\mathrm{T}}$.

System (23) is used here as a test problem for measuring the accuracy and efficiency of different τ-leap methods. We simulate the evolution of this system in the time interval [0,40] with initial condition $\mathbf{x}(0)=(10^4,0,0)^{\mathrm{T}}$ and reaction rates $\mathbf{c}=(0.1,0.002,0.5,0.04)^{\mathrm{T}}$. Figure 2 gives a simulation of this system obtained by the SSA.

We first give a brief description of the numerical process of the binomial τ-leap method. Similar numerical procedure can be obtained for the binomial midpoint τ-leap method. After choosing a stepsize τ by using the improved τ-selection process (2), the following process is used to generate binomial samples at each step.

(1) Check the Binomial stepsize conditions

$$R_1 \quad \text{and} \quad R_2, \quad \left[c_1x_1+\frac{1}{2}c_2x_1(x_1-1)\right]\tau/\left\lfloor\frac{1}{2}x_1\right\rfloor\leqslant 1;$$

$$R_3 \quad \text{and} \quad R_4, \quad (c_3+c_4)\tau\leqslant 1.$$

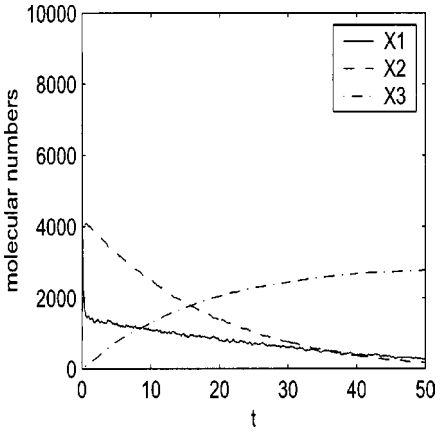

FIG. 2. A simulation of system 2 by the SSA.

(2) For reaction channels R_1 and R_2, generate a sample value K_{12} from the binomial random variable $\mathcal{B}\{\lfloor x_{1/2}\rfloor,[c_1x_1+c_2x_1(x_1-1)/2]\tau/[1/2x_1]\}$ and a sample value K_1 from $\mathcal{B}\{K_{12},c_1/[c_1+c_2(x_1-1)/2]\}$ for the reaction number of R_1. The reaction number of R_2 is $K_2=K_{12}-K_1$.

(3) For R_3 and R_4, generate a sample value K_{34} from the binomial random variable $\mathcal{B}[x_2,(c_3+c_4)\tau]$ and a sample value K_3 from $\mathcal{B}[K_{34},c_3/(c_3+c_4)]$ for the reaction number of R_3. The reaction number of R_4 is $K_4=K_{34}-K_3$.

(4) Update the system by $\mathbf{x}(t+\tau)=\mathbf{x}(t)+\Sigma_{j=1}^{4}\nu_jK_j$.

This system was simulated by the SSA, the Poisson and binomial τ-leap methods, and the binomial midpoint τ-leap method. We use the improved τ-selection process (2) to choose the stepsize with one of the three error control parameters: $\epsilon=0.05$, 0.03, and 0.01. For these three error control parameters, all of the stepsizes satisfy $\tau a_0(\mathbf{x})\geqslant 5$ and it is not necessary to use the SSA in the leap methods. This is a very good test system to test the accuracy and efficiency of the three leap methods. Programs were written in FORTRAN and computations were carried out in a Sun workstation with a 500 MHz CPU.

Regarding the SSA as giving exact results, we calculated the means and variances of molecular numbers at integer time points based on 20 000 simulations. For each method, we sum the absolute errors of the three molecular species in the mean and variance, that are presented in Fig. 3. The Poisson τ-leap method has slightly better accuracy in the first moment than the binomial τ-leap method. However, the accuracy in variance of the binomial τ-leap method is better than the Poisson τ-leap method. For a small error control parameter $\epsilon=0.01$, these two τ-leap methods have similar accuracy in the mean and variance. These results are consistent with those presented in Fig. 1. The midpoint τ-leap method can always achieve better accuracy of the first moment than the Poisson or binomial τ-leap method. An unexpected result is that the binomial τ-leap method has better accuracy in variance than the binomial midpoint τ-leap method.

Based on the computing time of the SSA, that is, 4188 s for 20 000 simulations, the speedup, defined by

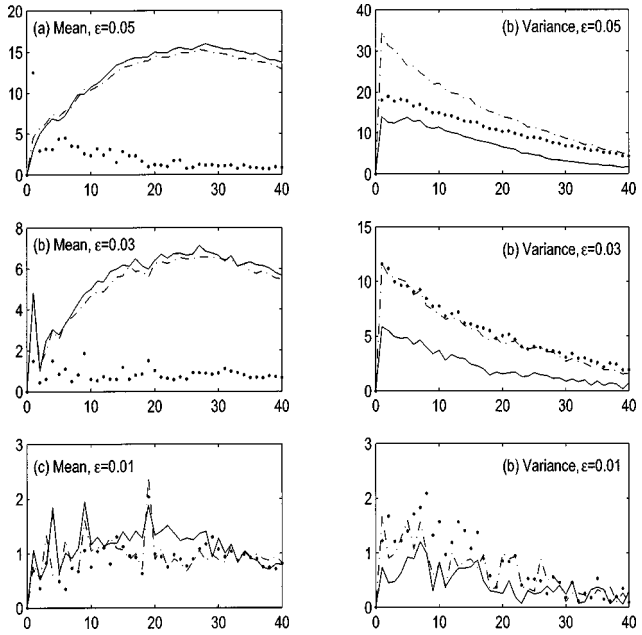

FIG. 3. Simulation results of the Poisson τ-leap method, Binomial τ-leap method and Binomial mid-point τ-leap method for system 2. [(a), (b) and (c)] Sum of absolute errors of the means of simulated molecular numbers of these three method with $\epsilon=0.05$, $\epsilon=0.03$, and $\epsilon=0.01$, respectively. [(d), (e), and (f)] Sum of absolute errors of the variances of simulated molecular numbers of these three method with $\epsilon=0.05$, $\epsilon=0.03$, and $\epsilon=0.01$, respectively. (Dash-line: the Poisson τ-leap method, solid line: the binomial τ-leap method, dot-line: the binomial midpoint τ-leap method).

$$\text{speedup} = \frac{\text{computing time of the SSA}}{\text{computing time of a } \tau\text{-leap method}} \quad (24)$$

is used to measure the improvement on efficiency of a τ-leap method over the SSA. Table I gives the averaged numbers of time steps of one simulation, computing time for 20 000 simulations (in seconds) and the speedup over the SSA. Based on the same error control parameter, these three τ-leap methods have similar computational efficiency in terms of the averaged number of time steps. The computational time of the Poisson τ-leap method is larger than that of the binomial τ-leap method due to different computational time for generating the random numbers. If the same time was used for generating Poisson and binomial random variables, the binomial τ-leap method would have slightly larger computing time than the Poisson τ-leap method. The computing time of the binomial midpoint τ-leap method is slightly larger than that of the binomial τ-leap method, since additional time is needed for midpoint prediction.

TABLE II. A full list of reaction channels and deterministic reaction rates for system 3.

	Reaction channel	Reaction rate
R_1	PLac+RNAP → PLacRNAP	0.17
R_2	PLacRNAP → PLac+RNAP	10
R_3	PLacRNAP → TrLacZ1	1
R_4	TrLacZ1 → RbsLacZ+PLac+TrLacZ2	1
R_5	TrLacZ2 → TrLacY1	0.015
R_6	TrLacY1 → RbsLacY+TrLacY2	1
R_7	TrLacY2 → RNAP	0.36
R_8	Ribosome+RbsLacZ → RbsRibosomeLacZ	0.17
R_9	Ribosome+RbsLacY → RbsRibosomeLacY	0.17
R_{10}	RbsRibosomeLacZ → Ribosome+RbsLacZ	0.45
R_{11}	RbsRibosomeLacY → Ribosome+RbsLacY	0.45
R_{12}	RbsRibosomeLacZ → TrRbsLacZ+RbsLacZ	0.4
R_{13}	RbsRibosomeLacY → TrRbsLacY+RbsLacY	0.4
R_{14}	TrRbsLacZ → LacZ	0.015
R_{15}	TrRbsLacY → LacY	0.036
R_{16}	LacZ → dgrLacZ	6.42E−5
R_{17}	LacY → dgrLacY	6.42E−5
R_{18}	RbsLacZ → dgrRbsLacZ	0.3
R_{19}	RbsLacY → dgrRbsLacY	0.3
R_{20}	LacZ+lactose → LacZlactose	9.52E−5
R_{21}	LacZlactose → product+LacZ	431
R_{22}	LacY → lactose+LacY	14

Simulations of this system suggest that random number generator is one of the key issues for the efficiency of the τ-leap methods. It is worthwhile to have a detailed study of the accuracy and efficiency of different random number generators for Poisson and binomial random variables. This issue is beyond the scope of this paper and will not be discussed here.

VI. SYSTEM 3: EXPRESSION AND ACTIVITY OF LacZ AND LacY

The third system describes the expression of LacZ and LacY genes and activity of LacZ and LacY proteins in *E. coli*. A detailed description of this system can be found in Kierzek.[20] Here we just give a full list of reaction channels and deterministic reaction rates of the chemical kinetics in Table II. As indicated by Kierzek,[20] reaction rates of the second-order reactions are dependent on the volume of cell.

This system was simulated by the software package STOCKS using the SSA.[20] Populations of the reactant species range from 0 or 1 for PLac to 30 000 for LacZ, and values of propensity functions range from 0.15 for reaction channel 5 (R_5), 24.0 for R_{12} to 500 000 for R_{20}. In addition, there is

TABLE I. Averaged numbers of time steps of one simulation, computing time for 20 000 simulations (in seconds), and the speedup over the SSA of the binomial midpoint τ-leap method, and binomial and Poisson τ-leap methods for simulating system 2.

	$\epsilon=0.05$			$\epsilon=0.03$			$\epsilon=0.01$		
	Steps	Time	Speedup	Steps	Time	Speedup	Steps	Time	Speedup
Poisson τ-leap	211	122	34.33	417	235	17.82	2902	1169	3.58
Binomial τ-leap	202	68	61.59	415	132	31.73	2902	793	5.28
Binomial midpoint	209	72	58.17	420	138	30.35	2904	827	5.06

TABLE III. Averaged computing time (in seconds) of one simulation of system 3 by using the SSA and binomial τ-leap method with different error control parameter ϵ in the improved τ-selection process.

Method	Computing time	Speedup
Binomial τ-leap (ϵ=0.05)	225.46	70.53
Binomial τ-leap (ϵ=0.03)	281.44	56.50
Binomial τ-leap (ϵ=0.01)	952.28	16.70
SSA	15 902	1

not any significant gap between the populations of different reactant species or between the values of propensity functions of different reaction channels. By using a multiscale method the improvement on efficiency is substantial but not significant.[13] Much time was used for carefully classifying reaction channels into three subsets of fast, intermediate, and slow reactions at each time step.

Now we discuss numerical implementation of the binomial τ-leap method for system 3. At each time step the numbers of RNAP and Ribosome are drawn from random pools that are distributed normally $N(35,3.5^2)$ and $N(350,35^2)$, respectively. In addition, the mean values of these pools grow, together with the volume of cell so that the concentrations of these molecules remain constant. For reaction channels R_1, R_8, and R_9, the numbers of trials of the binomial random variables are the populations of PLac, RbsLacZ, and RbsLacY, respectively.

For reaction channel R_{20}, the number of trials of the binomial random variable is the smaller of the populations of LacZ and Lactose. Denote the number of molecule P as $N(P)$, then the binomial random variable for R_{20} is defined by

$$\mathcal{B}(\min\{N(\mathrm{LacZ}),N(\mathrm{Lactose})\},\tau c_{20}\max\{N(\mathrm{LacZ}),N(\mathrm{Lactose})\}).$$

In addition, there are seven reactant species that undergo two reaction channels. These reactant species are PLacRNAP (channels R_2 and R_3), RbsLacZ (R_8 and R_{18}), RbsLacY (R_9 and R_{19}), RbsRibosomeLacZ (R_{10} and R_{12}), RbsRibosomeLacY (R_{11} and R_{13}), LacZ (R_{16} and R_{20}), and LacY (R_{17} and R_{22}). The simultaneous reaction sampling technique is applied to these pairs of reaction channels.

This system was simulated by the SSA and binomial τ-leap method with the improved τ-selection process (2). Table III gives the computational time for one simulation by using the SSA, which is averaged over 20 simulations, and by using the binomial τ-leap method with different error control parameter ϵ, which is averaged over 50 simulations. It took about 4.5 h for one simulation by using the SSA but much less time by using the binomial τ-leap method. The binomial τ-leap method with ϵ=0.03 results in a nearly 60-fold reduction in computational time over the SSA. We can still get a 16-fold improvement in computing time when a smaller ϵ=0.01 was used. The improvement in efficiency is significant.

Due to the huge computing time of the SSA, it is difficult to test the accuracy of the binomial τ-leap method based on a large number of simulations over the time interval [0,2100]. Instead we considered the accuracy of the binomial τ-leap method in a short time interval. As the SSA was frequently used at the initial stage of each simulation, we simulated this system in the time interval [300,330] and used the SSA to get a state of the system at $t=300$ that was used as the initial value of our simulations. We calculated the means and standard deviations of simulated molecular numbers over 10 000 simulations obtained by the SSA and binomial τ-leap method with error control parameter ϵ=0.03. As an example, we give the means and standard deviations of RbsLacY, TrLacZ2, TrRbsLacY, and TrRbsLacZ in Fig. 4. Simulation results suggest that the binomial τ-leap method can give very good approximations to the evolution of this biochemical reaction system.

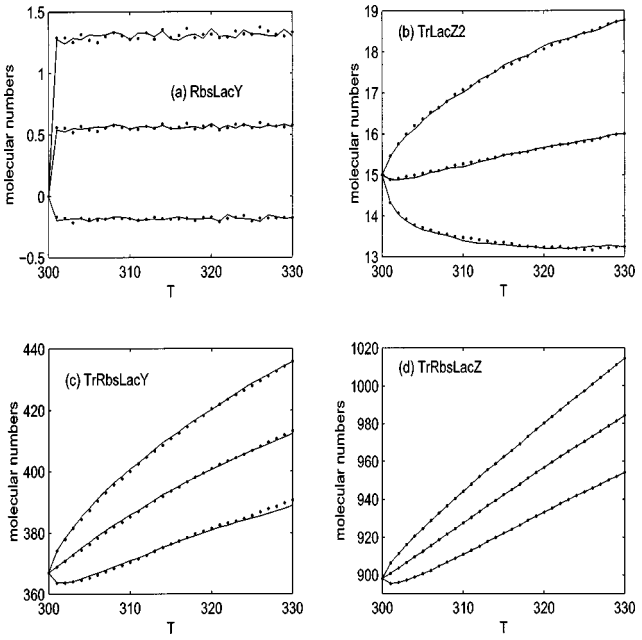

FIG. 4. Means and standard deviations ($\pm\sigma$) of four molecular species in system 3 by the SSA and binomial τ-leap method with ϵ=0.03 in the leap condition. (Solid line: the SSA, dot-line: the τ-leap method).

We note that it is very difficult to apply the Poisson τ-leap method to simulate this system. We simulated system 3 by using the Poisson τ-leap method with $k=3$ and 10 in the method selection criterion (3), respectively, and the SSA was used in k steps if $\tau a_0(\mathbf{x})<k$. All 100 simulations were aborted for each k due to negative molecular numbers. As the SSA was used frequently at the initial stage of each simulation, a larger k just delayed the time for using the Poisson τ-leap method and the time of abortion. Negative molecular number in most simulations was obtained from reaction channel R_8 or R_9 when the reaction number was 2 but the number of RbsLacZ or RbsLacY was just 1. The difficulty is that we may get negative number from channel R_8 or R_9 at any time point of a simulation. The numbers of RBsLacZ and RbsLacY are 0 or 1 at most steps when $0\leqslant t\leqslant 2100$ but the number of Ribosome is a sample from $N[350(1+t/2100),35^2]$. On the other hand, the sum of the expected reaction numbers of channels R_{20}, R_{21}, and R_{22} began to increase and exceeded k when t was large.

It is also difficult to apply the binomial midpoint τ-leap

method to simulate this system. The problem now is the midpoint prediction because the populations of a few reactant species are zero in the simulations. There are a number of reactant species, for example, PLac, PLacRNAP, TrLacZ1, and RbsLacZ, whose population is just 0, 1, or 2. If the molecular number is zero, any change made to it may cause unreasonable sample values for the reaction numbers.

VII. CONCLUSIONS

In this paper we have derived efficient numerical methods with robust leap control strategies for simulating chemical reaction systems. The motivation of this approach is to improve the efficiency of the Poisson τ-leap methods by using larger stepsizes and avoiding possible negative molecular numbers in the stochastic simulations. The first contribution of this approach is to use binomial random variables in the τ-leap methods. For a single reaction channel, the generated binomial sample, that represents the reaction number of this reaction channel in a given time interval, is less than or equal to the populations of reactant species that undergo this reaction. For simultaneous occurrence of different reaction channels, a sampling technique has been proposed for the total reaction number of a reactant species that undergoes two or more reaction channels. Numerical results for all three test systems indicated that the binomial leap methods can be used to simulate a wide range of chemical reaction systems with very good accuracy and significant improvement of efficiency over existing approaches. More work is needed in designing robust prediction strategies in the midpoint τ-leap method.

Another significant contribution of this paper is a robust leap control strategy. For a binomial random variable $\mathcal{B}(N,p)$, the probability p must satisfy $0 \leqslant p \leqslant 1$. This property has been used in the binomial leap methods as stepsize conditions to restrict the possible reaction numbers in a chosen time interval. If the leap condition is the leap control strategy based on the property of propensity functions, these stepsize conditions can be regarded as the leap control strategy based on the relationship of molecular numbers and reaction numbers. The combination of these two conditions provides a robust control strategy for practical applications.

Stepsize conditions can be used as additional conditions for choosing the leaping time. For the Lotka reactions,[6] for example,

$$\begin{aligned} R_1: &\quad S_1 + S_2 \xrightarrow{c_1} 2S_2, \\ R_2: &\quad S_2 + S_3 \xrightarrow{c_2} 2S_3, \\ R_3: &\quad S_3 \xrightarrow{c_3} S_4, \end{aligned} \tag{25}$$

it is not appropriate to generate sample values for the total reaction numbers of S_2 in channels R_1 and R_2 and S_3 in R_2 and R_3 at the same time. But the stepsize conditions of S_2 and S_3 can be used as additional conditions to choose the leaping size. This consideration for a simple system can be applied to more complex chemical reaction systems.

ACKNOWLEDGMENT

One of the authors (K.B.) would like to acknowledge funding from the Australian Research Council under the Federation Fellowship Program.

[1] A. Arkin, J. Ross, and H. H. McAdams, Genetics **149**, 1633 (1998).
[2] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, Science **422**, 633 (2003).
[3] D. Endy and R. Brent, Nature (London) **409**, 391 (2001).
[4] D. Gonze, J. Halloy, and A. Goldbeter, Proc. Natl. Acad. Sci. U.S.A. **99**, 673 (2002).
[5] J. Hasty, D. McMillen, F. Isaacs, and J. J. Collins, Nat. Rev. Genet. **2**, 268 (2001).
[6] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[7] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[8] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, J. Chem. Phys. **119**, 12784 (2003).
[9] K. Burrage and T. Tian, Advances in Scientific Computings and Applications, edited by Y Lu, W. Sun. and T. Tang (Science Press, Beijing/New York, 2004), p. 82.
[10] D. T. Gillespie and L. R. Petzold, J. Chem. Phys. **119**, 8229 (2003).
[11] C. Rao and A. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[12] E. L. Haseltine and J. B. Rawlings, J. Chem. Phys. **117**, 6959 (2002).
[13] K. Burrage, T. Tian, and P. M. Burrage, Prog. Biophys. Mol. Biol. **85**, 217 (2004).
[14] M. A. Gibson and J. Bruck, J. Phys. Chem. A **104**, 1876 (2000).
[15] D. T. Gillespie, J. Chem. Phys. **113**, 297 (2000).
[16] V. H. Resat, H. S. Wiley, and D. A. Dixon, J. Phys. Chem. B **105**, 11026 (2001).
[17] W. H. Press, B. Flannery, S. Teukosky, and W. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992).
[18] V. Kachitvichyanukul and B. M. Schmeiser, Commun. ACM **31**, 216 (1988).
[19] J. H. Ahrens and U. Dieter, ACM Trans. Math. Softw. **8**, 163 (1982).
[20] A. M. Kierzek, Bioinformatics **18**, 470 (2002).