Journal of Theoretical Biology 396 (2016) 90–104

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Modelling biochemical reaction systems by stochastic differential equations with reflection

Yuanling Niu [a,b,*], Kevin Burrage [c,d], Luonan Chen [b,**]

[a] School of Mathematics and Statistics, Central South University, Changsha 410083, China
[b] Key Laboratory of Systems Biology, Innovation Center for Cell Signaling Network, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Science, Chinese Academy of Sciences, China
[c] Department of Computer Science, Oxford University, Wolfson Building, Parks Road, Oxford OX1 3QD, UK
[d] School of Mathematical Sciences, Queensland University of Technology, Brisbane QLD 4001, Australia

HIGHLIGHTS

- A efficient approach for simulating biochemical reaction systems is provided.
- SDEs with reflection were employed to model the biochemical reaction systems.
- The domain where the species numbers should lie in is obtained.
- The projection in the numerical methods is actually a convex programming problem.
- The approach allows for incorporation in complex biological models in a tractable way.

ARTICLE INFO

Article history:
Received 27 June 2015
Received in revised form
17 January 2016
Accepted 6 February 2016
Available online 23 February 2016

Keywords:
Biochemical reaction system
Stochastic modelling
Stochastic simulation
Reflected stochastic differential equations
Reflection and correction approach

ABSTRACT

In this paper, we gave a new framework for modelling and simulating biochemical reaction systems by stochastic differential equations with reflection not in a heuristic way but in a mathematical way. The model is computationally efficient compared with the discrete-state Markov chain approach, and it ensures that both analytic and numerical solutions remain in a biologically plausible region. Specifically, our model mathematically ensures that species numbers lie in the domain D, which is a physical constraint for biochemical reactions, in contrast to the previous models. The domain D is actually obtained according to the structure of the corresponding chemical Langevin equations, i.e., the boundary is inherent in the biochemical reaction system. A variant of projection method was employed to solve the reflected stochastic differential equation model, and it includes three simple steps, i.e., Euler–Maruyama method was applied to the equations first, and then check whether or not the point lies within the domain D, and if not perform an orthogonal projection. It is found that the projection onto the closure $\overline{D}$ is the solution to a convex quadratic programming problem. Thus, existing methods for the convex quadratic programming problem can be employed for the orthogonal projection map. Numerical tests on several important problems in biological systems confirmed the efficiency and accuracy of this approach.

1. Introduction

Cells are intrinsically noisy biochemical reactors in which low reactant concentrations can lead to significant statistical fluctuations in molecule numbers and reaction rates. Most biological reactions occurring in a cell are inherently stochastic. Modeling and analysis of biochemical kinetics are important since they can provide insights into complicated systems where traditional experimentation is costly or difficult. Simulation is a powerful analysis technique, and in particular, stochastic simulation of biochemical systems has received increasing attention recently (Gillespie, 1977; Li, 2007; Székely et al., 2014; Niu et al., 2014; Székely, 2014).

The Chemical master equation (CME) takes a 'distribution' view of the biochemical reaction system, keeping track of every possible state. It describes the probabilities of the system being in each state at different time points. Therefore, the CME seems to be the ideal way of finding the evolution in time of a stochastic system, i.e., it is exact and returns the full probability density function

* Co-corresponding author at: School of Mathematics and Statistics, Central South University, Changsha 410083, China.
** Corresponding author.
E-mail addresses: yuanlingniu@csu.edu.cn (Y. Niu), lnchen@sibs.ac.cn (L. Chen).

http://dx.doi.org/10.1016/j.jtbi.2016.02.010
0022-5193/

(PDF) of the system. However, only in the case of very simple systems can the CME be solved analytically. In other words, for most cases, a numerical method is needed. Hence, solutions of the CME are generally not actually exact in practice. In addition, each solution is very computationally-intensive since each possible state has to be evaluated at each time point and the number of states can grow exponentially. Thus the CME approach could only be effectively used to simulate small-scale or simple systems for short periods of time initially. There has been considerable progress on CME solver methods in recent years (for example, Munsky and Khammash, 2006; MacNamara et al., 2008; MacNamara and Burrage, 2010). However, when populations and propensities grow too large to be calculated using the CME, trajectory-based simulations must be used instead.

The basic trajectory-based approach that exactly simulates individual paths from the full distribution given by the CME is the stochastic simulation algorithm (SSA) (Gillespie, 1977). We call it the discrete-state Markov chain approach. The SSA is a statistically exact method for generating Monte Carlo paths. A PDF built up from an infinite number of simulations of the SSA will be identical to the true distribution of the system, as given by the CME. Obviously this limit requiring infinite simulations cannot be reached, however a satisfactorily accurate PDF can be achieved by a moderately high number of repeats of the SSA. Although in recent years there has been considerable progress on Gillespie-type methods (Yates and Klingbeil, 2013; Anderson et al., 2014), these algorithms often become computationally expensive as increasingly complex physical systems are modelled, especially when there are wide ranges of rate constants and numbers of molecules.

The SDE model poses as an attractive alternative to the discrete-state Markov chain approach for studying the effects of stochastic behavior on the dynamics of excitable cells and tissues, due to the appreciable computational speed up achieved. However, solutions to the SDE model can become negative or even imaginary, and so have no physical meaning in such cases (e.g., the number of a molecule may become negative, violating its physical constraint). The SDE model breaks down in such cases. In the numerical algorithm, it is common to set the approximation to 0 when it becomes negative or resample the Wiener increment until the numerical solution is positive. We adopt the former approach in Section 3 when modelling biochemical reaction systems by SDE to avoid breaking down, and call it the modified SDE approach. But these alterations can bring in bias into the numerical solution as shown in Section 3. The authors (Wilkie and Wong, 2008) explored a way of modifying the noise term of the chemical Langevin equations (CLE) such that the variables preserve positivity. They allowed stochasticity only for those reaction steps in which a species participates. But the modified positivity preserving CLE system obtained in Wilkie and Wong (2008) after dropping the product noise terms does not guarantee the co-variance matrix matching condition and is thus physically inconsistent. The authors (Dana and Raha, 2011) studied a non-negativity preserving Milstein method implicit in both drift and diffusion terms. The scheme imposed the non-negativity constrained in the Newton iterations. Additionally, the scheme always satisfies the covariance matrix matching condition and is thus physically consistent. But the computational cost can be extremely expensive since a fully implicit method requires extra computation such as the Newton iteration at each time step. The authors (Schnoerr et al., 2014) show that, by extending the domain of the CLE to complex space, the CLE's accuracy for unimolecular systems is restored. It is shown that the complex CLE predicts real-valued quantities for the mean concentrations. But the molecule numbers are generally complex and are biologically unrealistic.

In this paper we address the issue of ensuring biologically plausible solutions to the SDE model without potentially biasing the solution compared with the modified SDE approach by incorporating boundaries into the SDE. The resulting model is called a SDE with reflection or reflected SDE and has previously been used to model human metabolic processes (Kawamura and Saisho, 2006), ion channel dynamics (Dangerfield et al., 2012; Dangerfield, 2012), and so on. The authors (Riley et al., 2008) developed a method for simulating Stochastic Hybrid Systems (SHS) with switching and reflecting boundaries. They pointed out that all biochemical systems cannot contain negative concentrations, and some biochemical processes also have saturation limits which impose upper limits on concentrations. This method enforces positivity of the molecule numbers by rejecting moves of the CLE algorithm which reduce the molecule numbers below zero. In addition, the modification may affect the accuracy of the CLE as an approximate method to probe stochastic chemical systems.

The field of SDE with boundary-condition model for biochemical reaction systems is much less well developed than for ordinary SDE model. SDE with reflection seems to be a good candidate for this problem. However, it is difficult to be carried out both analytically and numerically. A number of researches worked on it in a heuristic way, see for example (Riley et al., 2008). In this paper, we modelled the biochemical reaction systems by the reflected SDEs in a mathematical way. We gave a more accurate domain where the species numbers should lie in according to the structure of the CLE. Numerical approaches are used to solve the SDEs with reflection. Generally, they can be broadly split into two categories, i.e., penalization methods and projection methods. A variant of projection method was employed in this paper since numerical solutions of penalization methods can leave the domain, even if the exact solution to the reflected SDE does not. The only new feature of the projection method compared with the standard numerical methods for SDEs is the projection onto a convex polyhedron, that is usually easy to carry out. It is found that this projection is by chance a convex quadratic programming problem in the present work. General classes of methods that can be used to solve convex quadratic programming problems can also be used to find the projection. The outline of this paper is as follows. First, we propose our new model for biochemical reaction systems and determine the biochemical realistic region D. Then we introduce a simple numerical scheme to solve the reflected SDE model, and the new algorithm can actually be straightforward to be implemented. Finally, numerical results on several important biological problems confirm the effectiveness of our method.

2. The reflected SDE model for biochemical reaction systems

2.1. The biochemical reaction systems and the Langevin equation

Assume that there is a well-stirred biochemical reaction system containing N molecular species $S_1, S_2, \ldots, S_N$. These species of molecules chemically interact through M reaction channels at a constant temperature inside a fixed volume Ω. Let $x_i(t)(i=1,\ldots,N)$ denote the number of S_i at time t. Then the dynamic state of this system is $X=(x_1(t),x_2(t),\ldots,x_N(t))^T$. Assume the initial state is $X(t_0)=X_0$. Define the propensity functions $a_j(X(t))(j=1,\ldots,M)$ such that $a_j(X(t))\,dt$ are the probability that one reaction R_j will occur inside Ω in the next infinitesimal time interval $[t,t+dt)$ given $X(t)=X$. $a(X)=(a_1(X),a_2(X),\ldots,a_M(X))$. A state change vector ν_j is defined to characterize reaction R_j. The $N\times M$ matrix ν is called the stoichiometric matrix. Its element ν_{ij} represents the change in the number of species S_i due to reaction R_j.