Journal of Computational Physics 326 (2016) 398–419

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Hybrid framework for the simulation of stochastic chemical kinetics

Andrew Duncan [a,*], Radek Erban [b], Konstantinos Zygalakis [c]

[a] Department of Mathematics, Imperial College, South Kensington Campus, London, SW7 2AZ, United Kingdom
[b] Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom
[c] School of Mathematics, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, United Kingdom

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form 9 August 2016
Accepted 24 August 2016
Available online 30 August 2016

Keywords:
Chemical master equation
Chemical Langevin equation
Jump-diffusion process
Hybrid scheme

ABSTRACT

Stochasticity plays a fundamental role in various biochemical processes, such as cell regulatory networks and enzyme cascades. Isothermal, well-mixed systems can be modelled as Markov processes, typically simulated using the Gillespie Stochastic Simulation Algorithm (SSA) [25]. While easy to implement and exact, the computational cost of using the Gillespie SSA to simulate such systems can become prohibitive as the frequency of reaction events increases. This has motivated numerous coarse-grained schemes, where the "fast" reactions are approximated either using Langevin dynamics or deterministically. While such approaches provide a good approximation when all reactants are abundant, the approximation breaks down when one or more species exist only in small concentrations and the fluctuations arising from the discrete nature of the reactions become significant. This is particularly problematic when using such methods to compute statistics of extinction times for chemical species, as well as simulating non-equilibrium systems such as cell-cycle models in which a single species can cycle between abundance and scarcity. In this paper, a hybrid jump-diffusion model for simulating well-mixed stochastic kinetics is derived. It acts as a bridge between the Gillespie SSA and the chemical Langevin equation. For low reactant reactions the underlying behaviour is purely discrete, while purely diffusive when the concentrations of all species are large, with the two different behaviours coexisting in the intermediate region. A bound on the weak error in the classical large volume scaling limit is obtained, and three different numerical discretisations of the jump-diffusion model are described. The benefits of such a formalism are illustrated using computational examples.

1. Introduction

Biochemical systems with small numbers of interacting components have increasingly been studied in the recent years. Examples include the phage λ lysis-lysogeny decision circuit [5], circadian rhythms [57] and cell cycle [39]. It is this small number of interacting components that makes the appropriate mathematical framework for describing these systems a stochastic one. In particular, the kinetics of the different species is accurately described, under appropriate assumptions, by a continuous-time discrete-space Markov chain. The theory of stochastic processes [23,56] allows the association of the

* Corresponding author.
E-mail addresses: a.duncan@imperial.ac.uk (A. Duncan), erban@maths.ox.ac.uk (R. Erban), k.zygalakis@ed.ac.uk (K. Zygalakis).

http://dx.doi.org/10.1016/j.jcp.2016.08.034
0021-9991/

Markov chain with an underlying master equation, which is a set of ordinary differential equations (ODEs), possible of infinite dimensions, that describe, at each point in time, the probability density of all the different possible states of the system. In the context of biochemical systems this equation is known as the *chemical master equation* (CME).

The high dimensionality of the CME makes it intractable to solve in practice. In particular, with the exception of some very simple chemical systems [36] analytic solutions of the CME are not available. One way to deal with this issue is to resort to stochastic simulation of the underlying Markov chain. The stochastic simulation algorithm (SSA) developed by Gillespie [25] exactly simulates trajectories of the CME as the system evolves in time. The main idea behind this algorithm is that, at each time point, one samples a waiting time to the next reaction from an appropriate exponential distribution, while another draw of a random variable is then used to decide which of the possible reactions will actually occur. For suitable classes of chemically reacting systems, one can sometimes use exact algorithms which, although equivalent to the Gillespie SSA are less computationally intensive. Examples include the Gibson–Bruck Next Reaction Method [24] and the Optimised Direct Method [9]. These algorithms can be further accelerated by using parallel computing, for example, on Graphics Processing Units [43,42].

All the methods described above can only go so far in terms of speeding up the simulations, since even with all the possible speed ups running the SSA can be computationally intensive for realistic problems. One approach to alleviate the computational cost is to employ different approximations on the level of the description of the chemical system. For example, in the limit of large molecular populations, the waiting time becomes, on average, very small and under the law of mass action the time evolution of the kinetics is described by a system of ODEs. This system is known as the *reaction rate equation* which describes, approximately, the time evolution of the mean of the evolving Markov chain. An intermediate regime between the SSA and the reaction rate equation is the one where stochasticity is still important, but there exist a sufficient number of molecules to describe the evolving kinetics by a continuous model. This regime is called the chemical Langevin equation (CLE) [56,27], which is an Itô stochastic differential equation (SDE) driven by a multidimensional Wiener process. In this case the corresponding master equation for the CLE is called the chemical Fokker–Planck equation (CFPE) which is a N-dimensional parabolic partial differential equation, where N is the number of the different chemical species present in the system.

The fact that stochasticity is still present in the description of the chemical system, combined with the fact that the underlying CFPE is more amenable to rigorous analysis than the CME, has made the CLE equation a very popular regime used in applications [51,17,41]. However, while there are benefits to working with the CLE/CFPE, this approximation is only valid in the limit of large system volume and provides poor approximations for systems possessing one or more chemical species with low copy numbers. Furthermore, unlike the SSA/CME which ensures that there is always a positive (or zero) number of molecules in the system, the CFPE and CLE can give rise to negative concentrations, so that the chemical species can attain negative copy numbers. As observed in [52], negative copy numbers can occur even for extremely simple chemical reactions such as $\emptyset \to A$. This can have serious mathematical implications, since the CFPE equation might break down completely, due to regions in which the diffusion tensor is no longer positive definite, which makes the underlying problem ill-posed. On the level of the CFPE, one way to deal with such positivity issues is to truncate the domain and artificially impose no flux-boundary conditions along the domain boundary [11,18,53,17,29,10], which will have a negligible effect on the solution when it is concentrated far away from the boundary. When all chemical species exist in sufficiently high concentration, Dirichlet boundary conditions can also be used if one solves the stationary CFPE as an eigenvalue problem [46]. However, as shown in [15], these artificial boundary conditions can result in significant approximation errors when the solution is concentrated near the boundary. Other alternatives have been proposed to overcome the behaviour of the CLE close to the boundary, either by suppressing reaction channels which may cause negativity near the boundary [13], or by extending the domain of the process to allow exploration in the complex domain [52]. In the latter approach the resulting process, called the Complex CLE will have a positive definite diffusion tensor for all time, thus avoiding such breakdowns entirely. However, this method does not accurately capture the CME behaviour near the boundary, and in areas where the CLE is a poor approximation to the CME, the corresponding Complex CLE will suffer equally.

These issues have motivated a number of hybrid schemes which have been obtained by treating only certain chemical species as continuous variables and the others as discrete [34,20,50]. By doing so, such schemes are able to benefit from the computational efficiency of continuum approximations while still taking into account discrete fluctuations when necessary. Typically such schemes involve partitioning the reactions into “fast” and “slow” reactions, with the fast reactions modelled using a continuum approximation (CLE or the reaction rate equation), while using Markov jump process to simulate the discrete reactions. Chemical species which are affected by fast reactions are then modelled as continuous variables while the others are kept discrete. Since the reaction rate depends on the state, it is possible that some fast reactions become slow and vice versa. This is typically accounted for by periodically repartitioning the reactions. Based on this approach, a number of hybrid models have been proposed, such as [31,12], which couple deterministic reaction-rate equations for the fast reactions with Markov jump dynamics for the slow, resulting in a piecewise-deterministic Markov process for the entire system. Error estimates for such systems, in the large volume limit, were carried out in [38]. Similar methods have been proposed, such as [32] and more recently [54]. Other hybrid schemes [33,49] also involve a similar partition into slow and fast species, however the evolution of the slow species is obtained by solving the CME directly, coupled to a number of reaction-rate equations for the fast reactions. The hybrid system is thus reduced to a system of ODEs. An error analysis of these schemes was carried out in [37].