Mathematical Biosciences 292 (2017) 67–75

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Stochastic simulation of biochemical reactions with partial-propensity and rejection-based approaches

Vo Hong Thanh

The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Piazza Manifattura 1, Rovereto 38068, Italy

ARTICLE INFO

Article history:
Received 23 June 2016
Revised 4 January 2017
Accepted 3 August 2017
Available online 4 August 2017

Keywords:
Computational biology
Stochastic simulation
Partial-propensity rejection-based stochastic simulation algorithm

ABSTRACT

We present in this paper a new exact algorithm for improving performance of exact stochastic simulation algorithm. The algorithm is developed on concepts of the partial-propensity and the rejection-based approaches. It factorizes the propensity bounds of reactions and groups factors by common reactant species for selecting next reaction firings. Our algorithm provides favorable computational advantages for simulating of biochemical reaction networks by reducing the cost for selecting the next reaction firing to scale with the number of chemical species and avoiding expensive propensity updates during the simulation. We present the details of our new algorithm and benchmark it on concrete biological models to demonstrate its applicability and efficiency.

1. Introduction

The stochastic chemical kinetics is a promising framework to cope with complexities and randomnesses in understanding biological systems [1–6]. A discrete population of each individual species is kept track. Interactions between species are represented by biochemical reactions. Each reaction has a probability to fire in the next infinitesimal time proportional to a *propensity*. The propensity of a reaction depends on population of species involved and reaction kinetics. The dynamical behavior of biochemical reactions is described by the chemical master equation (CME) [7] and its solution can be realized by an exact simulation procedure called the stochastic simulation algorithm (SSA) [8,9]. SSA is a Monte-Carlo simulation technique that randomly selects a reaction to fire and to move the system to a new state according to a probability distribution derived under the hypothesis of CME.

The direct method (DM) and the first reaction method (FRM) [8] are the two well-known implementations of SSA. They are mathematically equivalent but differ in how to select the next reaction firing. FRM selects the reaction having the smallest putative time as the next reaction firing, while DM discovers the next reaction firing through a search. These basic algorithms, however, become computationally infeasible for practical models and many efficient formulations of these algorithms are introduced to enhance their efficiency. The optimized direct method (ODM) [10] and the sorting direct method (SDM) [11] improve the search for next reaction firings of DM by sorting reactions based on their firing frequencies. The multi-dimensional search is an attempt to reduce the computational time complexity of the search for next reaction firings by dividing reactions into groups [12]. The selection of reaction firing in the multi-dimensional search is composed of two steps: 1) selecting the group, then 2) locating the next reaction within that group. The finest strategy for grouping of reactions is when each group contains only two reactions which is equivalent to a tree structure where reactions are stored on its leaves. The search of the next reaction firing in this case is a tree traversal procedure [13–16]. The SSA with composition-rejection search strategy (SSA-CR) [17,18] also groups reaction into groups, but the selection of the next reaction firing in a group employs a rejection-based sampling instead. The next reaction method (NRM) [19] is an alternative which focuses on improving FRM. It uses a binary heap to store and extract smallest (absolute) reaction firing times. Other improvements including approximate and parallel algorithms [20–25] are also introduced.

The partial-propensity direct method (PDM) [26–29] and the rejection-based SSA (RSSA) [30–34] are two exact simulation approaches that have been introduced recently to improve the performance of the stochastic simulation. They focus on different simulation bottlenecks of SSA to improve the simulation performance. The former improves the search for next reaction firings, while the latter reduces the propensity updates after reaction firings. PDM introduces the concept of partial propensity function to factorize propensities of reactions. The factors are grouped by common reactants in a so-called *partial propensity structure*. The search for the next reaction firing using the partial propensity data structure is done by first sampling the group index and then the element index

E-mail address: vo@cosbi.eu

http://dx.doi.org/10.1016/j.mbs.2017.08.001
0025-5564/

at which the partial propensity of the reaction is stored. The variants of PDM including the sorting PDM (SPDM) [26], the PDM with composition-rejection search (PSSA-CR) [27] are also introduced to improve the selection of the next reaction firing. The current limitation of PDM is that it is only applicable to the class of reactions involving at most two reactants (i.e., *elementary reactions*) and their propensities must be factorizable (i.e., *mass-action* propensity) [26]. The network that contains non-elementary reactions must be decomposed into elementary ones at the expense of increasing network size. On the other side, RSSA focuses on reducing the average number of propensity updates during the simulation. The propensity of each reaction in RSSA is abstracted into an interval denoted by a pair of propensity lower bound and upper bound. The propensity bounds of reactions are derived by specifying an arbitrary bound on the population of each species, which is called *fluctuation interval* (or *abstract state*). RSSA selects the next reaction firings using the propensity bounds in two steps. First, a candidate reaction is randomly selected proportionally to its propensity upper bound. Then, a rejection-based test on the candidate is performed to ensure that it fires with the same probability determined by SSA. After the state is updated by the reaction firing, a new selection is performed without recomputing the propensity bounds except in uncommon cases where the population of a species exits its fluctuation interval.

We present in this paper a new exact simulation algorithm, which is called the partial-propensity rejection-based SSA (PRSSA), to offer the computational advantages of the partial-propensity and rejection-based approaches. Our algorithm applies the principle of the rejection-based approach on the concept of partial propensity to define partial propensity bounds for reactions, then factorizes the propensity bounds of reactions by using the interval analysis [35]. Based on such factorization, PRSSA groups the partial propensity bounds of reactions having sharing reactants. The grouping of partial propensity bounds is then used by PRSSA to select the next reaction firing. By employing partial propensity bounds of reactions for the selection of next reaction firings, PRSSA reduces its computational cost for selection of the next reaction firing to scale with the number of species while skipping many expensive propensity updates during the simulation. Our new algorithm provides a favorable computational complexity for the simulation of reaction networks where the number of reactions is much larger than the number of species and the number of reactions that need updates their propensities when a reaction firing is large.

The paper is organized as follows. Section 2 provides the background of stochastic simulation of biochemical reactions. Section 3 presents our new PRSSA algorithm. We describe in detail how to combine rejection-based approach and the concept of partial-propensity function to select the next reaction firing in order to reduce the computation cost of SSA. Section 4 shows the numerical results of our algorithm on concrete models acting as benchmarks to demonstrate the applicability and efficiency with respect to the state of the art algorithms. The concluding remarks are in Section 5.

2. Stochastic simulation

We consider a well-mixed biochemical reaction network consisting of N molecular species S_i for $i = 1 \dots N$. Let $X_i(t)$ be the absolute number denoting the population of species S_i at time t. The state of the system at the time t is a N-vector $X(t) = (X_1(t), \dots, X_N(t))$ that shows the population of each species in the system at the time.

Species can interact with each other through M reactions R_j for $j = 1 \dots M$. A reaction R_j between species has a general form:

$$R_j : \nu_{1j}^- S_1 + \dots + \nu_{nj}^- S_n \xrightarrow{c_j} \nu_{1j}^+ S_1 + \dots + \nu_{nj}^+ S_n \quad (1)$$

where c_j is the *stochastic reaction constant* [1,9,36]. The species on the left side of the arrow are *reactants*, while the ones on the right side are *products*. The non-negative integer ν_{ij}^- and ν_{ij}^+ called *stoichiometric coefficients* give the number of molecules a reactant consumed and the number of molecules a product produced. The N-vector ν_j, where the ith element is $\nu_{ij} = \nu_{ij}^+ - \nu_{ij}^-$ represents the changes in the population species S_i due to the firing of R_j, is the state change vector. The vector ν_j denotes the net change in population of each species when firing R_j. Formally, if reaction R_j fires at time $t + \tau$, given the state $X(t)$ at time t, then the system jumps to a new state $X(t + \tau) = X(t) + \nu_j$.

Each reaction R_j has a probability to fire in the next infinitesimal time that is proportional to a *propensity* a_j. The propensity function a_j is defined so that $a_j(X(t))dt$ gives the probability that the reaction R_j fires in the next infinitesimal time $t + dt$ given the system state $X(t)$ at time t. An explicit formula of the propensity of a reaction on the state is depending on the chemical kinetics. For the mass-action kinetics, propensity a_j of a reaction R_j is defined as:

$$a_j(X(t)) = c_j h_j(X(t)) \quad (2)$$

where $h_j(X(t))$ counts the number of distinct combinations of reactants involved in R_j and c_j is its stochastic reaction constant. The number of combinations of reactants of a *synthesis reaction*, which is used to introduce new molecular species into the system from an external source, is set $h_j(X(t)) = 1$.

The mathematical framework of the stochastic simulation is the joint probability density function (pdf) $p(\tau, \mu)$ which is defined such that $p(\tau, \mu)d\tau$ gives the probability that a reaction fires in the next infinitesimal time interval $[t + \tau, t + \tau + d\tau)$ and it is reaction $R\mu$, given the state $X(t)$ at time t. The analytical formula of $p(\tau, \mu)$ is given as:

$$p(\tau, \mu) = a_\mu exp(-a_0 \tau) \quad (3)$$

where $a_0 = \sum_{j=1}^{M} a_j$.

The stochastic simulation algorithm (SSA) is a class of exact algorithm for sampling the pdf $p(\tau, \mu)$ in Eq. (3) by using an observation that the probability that reaction R_μ occurs in the next time $t + \tau$ follows a discrete probability distribution a_μ/a_0 and the firing time τ is an exponential distribution $\mathrm{Exp}(a_0)$. So, to construct a simulation trajectory of the reaction network, SSA repeatedly selects the next reaction firing R_μ with probability a_μ/a_0 and generates its firing time τ from the exponential distribution $\mathrm{Exp}(a_0)$. It then advances the time by an amount τ and updates the state accordingly to the selected reaction R_μ. The propensities of reactions are updated as well to reflect changes in the system state. These simulation steps are repeated until a predefined ending time is reached.

3. Partial-propensity rejection-based SSA

This section presents the partial-propensity rejection-based SSA (PRSSA) for improving the performance of exact stochastic simulation of biochemical reactions. We first cover the backgrounds on the rejection-based and partial-propensity approaches. Then, we use these concepts to derive and group the partial propensity bounds of reactions. We present the data structures and the detailed implementations of our new algorithm. Employing such data structures, the selection of reaction firings in PRSSA is scaled with the number of species while propensity updates are avoided during the simulation.

3.1. Background on rejection-based SSA

The rejection-based stochastic simulation algorithm (RSSA) [31] is an exact simulation with the aim to reduce the average