BioSystems 149 (2016) 26–33

Contents lists available at ScienceDirect

BioSystems

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

Stochastic analysis of Chemical Reaction Networks using Linear Noise Approximation☆

Luca Cardelli[a,b], Marta Kwiatkowska[a], Luca Laurenti[a,*]

[a] Department of Computer Science, University of Oxford, United Kingdom
[b] Microsoft Research, Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 27 November 2015
Received in revised form 8 July 2016
Accepted 1 September 2016
Available online 29 October 2016

Keywords:
Chemical Reaction Networks
Linear Noise Approximation
Probabilistic logic
Model checking

ABSTRACT

Stochastic evolution of Chemical Reactions Networks (CRNs) over time is usually analyzed through solving the Chemical Master Equation (CME) or performing extensive simulations. Analysing stochasticity is often needed, particularly when some molecules occur in low numbers. Unfortunately, both approaches become infeasible if the system is complex and/or it cannot be ensured that initial populations are small. We develop a probabilistic logic for CRNs that enables stochastic analysis of the evolution of populations of molecular species. We present an approximate model checking algorithm based on the Linear Noise Approximation (LNA) of the CME, whose computational complexity is independent of the population size of each species and polynomial in the number of different species. The algorithm requires the solution of first order polynomial differential equations. We prove that our approach is valid for any CRN close enough to the thermodynamical limit. However, we show on four case studies that it can still provide good approximation even for low molecule counts. Our approach enables rigorous analysis of CRNs that are not analyzable by solving the CME, but are far from the deterministic limit. Moreover, it can be used for a fast approximate stochastic characterization of a CRN.

1. Introduction

Chemical Reaction Networks (CRNs) and mass action kinetics are well studied formalisms for modelling biochemical systems (Chellaboina et al., 2009). In recent years, CRNs have also been successfully used as a formal programming language for biochemical systems (Soloveichik et al., 2010; Cardelli, 2013; Chen et al., 2013). There are two well established approaches for analysing chemical networks: deterministic and stochastic (Gillespie et al., 2013). The deterministic approach models the kinetics of a CRN as a system of ordinary differential equations (ODEs) and represents average behaviour, valid in the thermodynamic limit, when the molecular population is sufficiently high (Gillespie, 2009). The stochastic approach, on the other hand, is based on the Chemical Master Equation (CME) and models the CRN as a continuous-time Markov chain (CTMC) (Cardelli, 2008). The stochastic behaviour can be analyzed by stochastic simulation (Gillespie et al., 2013) or by exhaustive probabilistic model checking of the CTMC, which can be performed, for example, by using PRISM (Kwiatkowska et al., 2011). Exhaustive analysis of the CTMC is able to find the best- and worst-case scenarios and is correct for any population size, but suffers from the state-space explosion problem (Kwiatkowska and Thachuk, 2014) and can only be used for relatively small systems. In contrast, deterministic methods are much more robust with respect to state-space explosion, but unable to represent stochastic fluctuations, which play a fundamental role when the system is not in thermodynamic equilibrium. As a consequence, approximate approaches to efficiently solve the CME are appealing. For instance, in Ammar et al. (2012) and Chinesta et al. (2015), the authors use proper generalized decomposition in order to efficiently derive a numerical solution of the CME. These approaches are based on the assumption that the probability of the system being in a particular state can be written as a finite sum of separable functions. Herein, we consider a different approach based on a continuous stochastic approximation of the CME (Van Kampen, 1992).

1.1. Contributions

In this paper we develop a novel approach for analysing the stochastic evolution of a CRN based on the Linear Noise

☆ Authors order is alphabetical.

* Corresponding author at: Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3BH, United Kingdom.
E-mail addresses: luca@microsoft.com (L. Cardelli), marta.kwiatkowska@cs.ox.ac.uk (M. Kwiatkowska), luca.laurenti@cs.ox.ac.uk (L. Laurenti).

http://dx.doi.org/10.1016/j.biosystems.2016.09.004

Approximation (LNA) of the CME. We formulate SEL (Stochastic Evolution Logic), a probabilistic logic for CRNs that enables reasoning about probability, expectation and variance of linear combinations of populations of the species. Examples of properties that can be specified in our logic include "the maximum expected population size of species λ_1 during the first 20 s is 75 molecules" and "the probability that the combined population of species λ_1 and λ_2 has degraded between 10 and 30 s is less than 0.1". We propose an approximate model checking algorithm for the logic based on the LNA and implement it in Matlab and Java. We demonstrate that the complexity of model checking is polynomial in the initial number of species and independent of the initial molecule counts, thus ameliorating state-space explosion. Further, we show that model checking is exact when approaching the thermodynamic limit. Though the algorithm may not be accurate for systems far from the deterministic limit, this generally happens when the populations are small, in which case the analysis can be performed by transient analysis of the induced CTMC (Kwiatkowska et al., 2007). Our approach is essential for CRNs that cannot be analyzed by (partial) state space exploration, because of large or infinite state spaces. Moreover, it is useful for a fast (approximate) stochastic characterization of CRNs, since solving the LNA is much faster than solving the CME (Elf and Ehrenberg, 2003). We prove asymptotic correctness of LNA-based model checking and show on four examples that it is still possible to obtain very good approximations even for small population systems, compared to standard uniformisation (Kwiatkowska et al., 2007) and statistical model checking implemented in PRISM (Kwiatkowska et al., 2011).

1.2. Related work

The closest work to ours is by Bortolussi and Lanciani (2013), which uses the Central Limit Approximation (CLA) (essentially the same as the LNA) for checking restricted timed automata specifications, assuming a fixed population size. Wolf et al. (2010) develop a sliding window method to approximately verify infinite-state CTMCs, which applies to cases where most of the probability mass is concentrated in a confined region of the state space. Recently, Finite State Projection algorithms (FSP algorithms) for the solution or approximation of the CME have been introduced (Munsky and Khammash, 2006). Sliding window and FSP algorithms apply to the induced CTMC, but require at least partial exploration of the state space, and are thus not immune to state-space explosion. Moment closure techniques (Singh and Hespanha, 2068; Hespanha, 2008) improve scalability by estimating the first $k \in \mathbb{N}$ moments of the distribution of the species over time. The LNA itself can be seen as a moment closure technique, as a Gaussian distribution is completely characterized by the first two moments. However, the LNA tells us more because it guarantees that, if certain conditions are satisfied, the distribution of the process is Gaussian.

Continuous Stochastic Logic (CSL), originally introduced in Aziz et al. (2000) and extended by Baier et al. (2003), is a logic widely used to perform model checking of continuous-time Markov chains. CSL combines temporal operators of the logic CTL with the probabilistic and steady-state operators, and is further extended with reward operators in Kwiatkowska et al. (2007). CSL model checking is based on solving the CME and proceeds through uniformisation of the CTMC, essentially a time discretisation, and thus involves traversal of the full state space. This can be partially ameliorated by fast adaptive uniformisation (Dannenberg et al., 2015) that does not consider states with negligible probability. An alternative is statistical model checking (SMC) which involves a key operator of CSL is probabilistic reachability, that is, computing the probability that a particular region of the state space is reached over a given time interval. Although SEL is endowed with a probabilistic operator, this operator gives the average value of the probability over time and, if the time interval is not a singleton, this is not equivalent to probabilistic reachability. Nevertheless, as shown in Bortolussi et al. (2016), SEL and our approximate model checking algorithm can be extended to express reachability, but currently lacks reward operators. The CSL steady-state operator of CSL cannot be added to CSL because LNA is accurate only for finite time. PRISM (Kwiatkowska et al., 2011) implements CSL model checking using uniformisation, fast adaptive uniformisation and statistical model checking.

Hybrid Automata Stochastic Logic HASL (Ballarini et al., 2011a) is an expressive specification formalism for stochastic Petri nets based on linear hybrid automata that is employed by the tool Cosmos (Ballarini et al., 2011b). CRNs have a natural interpretation in terms of stochastic Petri nets, see e.g. Barbot and Kwiatkowska (2015). The HASL formalism is more expressive than SEL and CSL, and can express CSL probabilistic reachability and expected reward properties. HASL model checking proceeds through statistical model checking of the product of a HASL specification automaton and the Petri net, and is implemented in Cosmos. In contrast, SEL model checking follows through approximating the solution of the CME with the Gaussian process induced by the LNA.

1.3. Structure of the paper

In Section 2 we summarize the deterministic and stochastic modelling approaches for CRNs, and in Section 3 we describe the Linear Noise Approximation method. Section 4 introduces the logic SEL and the corresponding model checking algorithm based on the LNA. In Section 5 we demonstrate our approach on four networks taken from the literature. Section 6 concludes the paper.

2. Chemical Reaction Networks

A *Chemical Reaction Network (CRN)* $C = (\Lambda, R)$ is a pair of finite sets, where Λ is the set of *chemical species* and R the set of reactions. $|\Lambda|$ denotes the size of the set of species. A *reaction* $\tau \in R$ is a triple $\tau = (r_\tau, p_\tau, k_\tau)$, where $r_\tau, p_\tau \in \mathbb{N}^{|\Lambda|}$ and $k_\tau \in \mathbb{R}_{>0}$. r_τ and p_τ represent the stoichiometry of reactants and products and k_τ is the coefficient associated to the rate of the reaction; its dimension is s^{-1}. We often write reactions as $\lambda_1 + \lambda_3 \to^{k_1} 2\lambda_2$ instead of $\tau_1 = ([1,0,1]^T, [0,2,0]^T, k_1)$, where $\cdot^T$ indicates the transpose of a vector. We define the *net change* associated to a reaction τ by $\upsilon_\tau = p_\tau - r_\tau$. For example, for τ_1 as above, we have $\upsilon_{\tau_1} = [-1, 2, -1]^T$.

We make the assumption that the system is well stirred, that is, the probability of the next reaction occurring between two molecules is independent of the location of those molecules. We consider fixed volume V and temperature; under these assumptions a *configuration* or *state* $x \in \mathbb{N}^{|\Lambda|}$ of the system is given by the number of molecules of each species. We define $[x] = x/N$, the vector of the species *concentration* in x for a given N, where $N = V \cdot N_A$ is the volumetric factor, V is the volume of the solution and N_A is Avogadro's number. The physical dimension of N is mol^{-1} L, where mol indicates mole and L is litre. Given $\lambda_i \in \Lambda$ then $\#\lambda_i\_x \in \mathbb{N}$ represents the number of molecules of λ_i in x and $[\lambda_i]\_x \in \mathbb{R}$ the concentration of λ_i in the same configuration. In some cases we elide x, and we simply write $\#\lambda_i$ and $[\lambda_i]$ instead of $\#\lambda_i\_x$ and $[\lambda_i]\_x$. They are related by $[\lambda_i] = \#\lambda_i/N$. The dimension of $[\lambda_i]$ is $\mathrm{mol\,L}^{-1}$.

The propensity $\alpha_{n,\tau}$ of a reaction τ in terms of the number of molecules (here subscript n stands for the number of molecules) is a function of the current configuration of the system x such that $\alpha_{n,\tau}(x)\mathrm{d}t$ is the probability that a reaction event occurs in the next infinitesimal interval $\mathrm{d}t$. In this paper we assume as valid the stochastic form of the law of mass action, so the propensity rates are proportional to the number of molecules that participate in the reaction (Cardelli, 2008). Stochastic models consider the system in terms of numbers of molecules, while deterministic ones,