

Exploring Parameter Space of Stochastic Biochemical Systems Using Quantitative Model Checking^{*}

Luboš Brim, Milan Češka, Sven Dražan, and David Šafránek

Systems Biology Laboratory at Faculty of Informatics, Masaryk University,
Botanická 68a, 602 00 Brno, Czech Republic
{brim,xceska,xdrazan,xsafran1}@fi.muni.cz

Abstract. We propose an automated method for exploring kinetic parameters of stochastic biochemical systems. The main question addressed is how the validity of an *a priori* given hypothesis expressed as a temporal logic property depends on kinetic parameters. Our aim is to compute a landscape function that, for each parameter point from the inspected parameter space, returns the quantitative model checking result for the respective continuous time Markov chain. Since the parameter space is in principle dense, it is infeasible to compute the landscape function directly. Hence, we design an effective method that iteratively approximates the lower and upper bounds of the landscape function with respect to a given accuracy. To this end, we modify the standard uniformization technique and introduce an iterative parameter space decomposition. We also demonstrate our approach on two biologically motivated case studies.

1 Introduction

The importance of stochasticity in biochemical processes having low numbers of molecules has resulted in the development of stochastic models [12]. Stochastic biochemical processes can be faithfully modeled as continuous time Markov chains (CTMCs) [9]. Knowledge of stochastic rate constants (model parameters) is important for the analysis of system dynamics. Moreover, knowledge about how change in parameters influences system dynamics (parameter exploration) is of great importance in tuning the stochastic model. Prior knowledge of kinetic parameters is usually limited. The model identification routine thus typically includes parameter estimation based on experimental data. While parameter exploration and estimation is well-established for deterministic models, it has not yet been adequately addressed and sufficiently developed for stochastic models. The purpose of this work is to develop practical and effective methods for exact exploration of model parameters in stochastic biochemical models.

The main question addressed is how the validity of an *a priori* given hypothesis expressed as a temporal property depends on model parameters. Parameter estimation

^{*} This work has been supported by the Czech Science Foundation grant No. GAP202/11/0312.

M. Češka has been supported by Ministry of Education, Youth, and Sport project No. CZ.1.07/2.3.00/30.0009 – Employment of Newly Graduated Doctors of Science for Scientific Excellence. D. Šafránek has been supported by EC OP project No. CZ.1.07/2.3.00/20.0256.

gives a single point in the parameter space where the values of model parameters maximize the agreement of model behaviour with experimental data. On the contrary, we often do not want to have a single objective but rather explore the property over the entire parameter space. Our main goal is to compute a *landscape function* that for each parameter point from the inspected parameter space returns the quantitative model checking result for the respective CTMC determined by the parameter point and the given property. Since the inspected parameter space is in principle dense the set of parametrized CTMCs to be explored is infinite. It is thus not possible to compute the model checking result for each CTMC individually.

As a temporal logic we use the *bounded time fragment of Continuous Stochastic Logic (CSL)* [2] further extended with rewards [19]. For most cases of biochemical stochastic systems the bounded time restriction is adequate since a typical behaviour is recognizable in finite time intervals.

In this paper we consider the *parameter exploration problem* for stochastic biochemical systems in terms of a landscape function that returns for each parameter point the probability or the expected reward of the inspected CSL formula. We propose a method, called *min-max approximation*, that computes the *lower* and *upper* approximations of the landscape function. To compute the approximation for an arbitrary nested CSL formula, we introduce the largest and smallest set of states satisfying the formula and show how to compute such sets effectively using a new method called *parametrized uniformization*. To compute the landscape function approximation with given accuracy we employ *iterative parameter space decomposition* that divides the parameter space into subspaces and allows to compute the proposed approximation independently for each subspace. This decomposition refines the approximation and enables to reach the required accuracy bound. We demonstrate our approach on two biologically motivated case studies. In the first one, we demonstrate that parametrized uniformization allows to approximate the transient probabilities of Schloegel’s model [24] for the inspected parameter space. In the second case, our method is applied to parameter exploration of bi-stability in mammalian cell cycle gene regulatory control [25]. Several techniques have been employed [26,10] to analyze models of this kind, especially, it has been shown that asymptotic solutions may disagree with the exact solution imposing thus a challenge for more accurate computational techniques. Since in low molecular numbers stochasticity can produce behaviour that significantly differs from asymptotic and deterministic dynamics, the parameter exploration method reflecting this phenomenon is very important for computational systems biology.

In contrast to methods mentioned in the related work section, the accuracy of results can be fully controlled and adjusted by the user. Similarly to these methods our method is computationally intensive. However, it can be easily parallelized since the computation for each subspace is independent. Moreover, it can be also combined with fast adaptive uniformization [9] and sliding window abstraction [16].

Related Work. To the best of our knowledge there is no other work on stochastic models employing CSL model checking to systematic parameter exploration. The closest work is [22] where a CTMC is explored with respect to a property formalized as a deterministic timed automaton (DTA). It extends [1] to parameter estimation with respect to acceptance of the DTA. Approaches to parameter estimation [23,1,7] rely on

approximating the maximum likelihood. Their advantage is the possibility to analyse infinite state spaces [1] (employing dynamic state space truncation with numerically computed likelihood) or even models with no prior knowledge of parameter ranges [7] (using Monte-Carlo optimization for computing the likelihood). All these methods are not suitable for computing the landscape function since they focus on optimizing a *single objective* and not on global *exploration* of the entire requested parameter space.

Approaches based on Markov Chain Monte-Carlo sampling and Bayesian inference [13,17,18] can be extended to sample-based approximation of the landscape function, but at the price of undesired inaccuracy and high computational demands [6,4]. Compared to these methods, our method provides an exact result without neglecting any singularities caused by possible discontinuities in the landscape function.

In [15], for a given parametrized discrete time Markov chain (DTMC) the problem of synthesis for a probabilistic temporal logic is considered. The problem is reduced to constructing a regular expression representing the property validity while addressing the problem of expression explosion. Construction of the expression and also the proposed reduction techniques rely on the discrete nature of DTMC. These techniques cannot be successfully applied to CTMC since the complexity of the expression is given by maximal number of events that can occur within the inspected time horizon. In a typical biochemical system where time scales of individual reactions differ in several orders the number of reactions that can occur is enormous.

Barbuti et al. [5] treat stochastic biochemical models with parameter uncertainty in terms of interval discrete time Markov chains. They reduce quantitative reachability analysis of uncertain models to reachability analysis of a Markov Decision Process. However, no analogy to landscape function and automatized parameter decomposition is considered. Moreover, our method deals with continuous time semantics.

2 Background

Stochastic Biochemical Systems. A finite state stochastic biochemical system \mathcal{S} is defined by a set of N *chemical species* in a well stirred volume with fixed size and fixed temperature participating in M *chemical reactions*. The number X_i of molecules of each species S_i has a specific bound and each reaction is of the form $u_1S_1 + \dots + u_NS_N \longrightarrow v_1S_1 + \dots + v_NS_N$ where $u_i, v_i \in \mathbb{N}_0$ represent *stoichiometric coefficients*.

A *state* of a system in time t is the vector $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_N(t))$. When a single reaction with index $r \in \{1, \dots, M\}$ with vectors of stoichiometric coefficients U_r and V_r occurs the state changes from \mathbf{X} to $\mathbf{X}' = \mathbf{X} - U_r + V_r$, which we denote as $\mathbf{X} \xrightarrow{r} \mathbf{X}'$. For such reaction to happen in a state \mathbf{X} all reactants have to be in sufficient numbers and the state \mathbf{X}' must reflect all species bounds. The *reachable state space* of \mathcal{S} , denoted as \mathbb{S} , is the set of all states reachable by a finite sequence of reactions from an *initial state* \mathbf{X}_0 . For each state \mathbf{X}_i we denote $\text{pred}(\mathbf{X}_i) = \{(j, r) \mid \mathbf{X}_j \xrightarrow{r} \mathbf{X}_i\}$ and $\text{succ}(\mathbf{X}_i) = \{(j, r) \mid \mathbf{X}_i \xrightarrow{r} \mathbf{X}_j\}$ the sets of all predecessors and successors, respectively, together with indices of corresponding reactions. The set of indices of all reactions changing the state \mathbf{X}_i to the state \mathbf{X}_j is denoted as $\text{reac}(\mathbf{X}_i, \mathbf{X}_j) = \{r \mid \mathbf{X}_i \xrightarrow{r} \mathbf{X}_j\}$. Henceforward the reactions will be referred directly by their indices.

According to Gillespie [12] the behaviour of a stochastic system \mathcal{S} can be described by the continuous time Markov chain (CTMC) $C = (\mathbb{S}, \mathbf{X}_0, \mathbf{R})$ where the transition matrix $\mathbf{R}(i, j)$ gives the probability of a transition from \mathbf{X}_i to \mathbf{X}_j . Formally, $\mathbf{R}(i, j) = \sum_{r \in \text{reac}(\mathbf{X}_i, \mathbf{X}_j)} k_r \cdot C_{r,i}$ such that k_r is a *stochastic rate constant* of the reaction r and $C_{r,i} \stackrel{\text{def}}{=} \prod_{l=1}^N \binom{\mathbf{X}_{i,l}}{u_l}$ corresponds to the population dependent term of the *propensity function* where $\mathbf{X}_{i,l}$ is l th component of the state \mathbf{X}_i and u_l is the stoichiometric coefficient of the reactant S_j in reaction r .

Parameter Space. Let each stochastic rate constant k_i have a value interval $[k_i^\perp, k_i^\top]$ with minimal and maximal bounds expressing *uncertainty range* of its value. A *parameter space* \mathbf{P} induced by a set of stochastic rate constants k_i is defined as the Cartesian product of the individual value intervals $\mathbf{P} = \prod_{i=1}^M [k_i^\perp, k_i^\top]$. A single *parameter point* $p \in \mathbf{P}$ is an M -tuple holding a single value of each rate constant $p = (k_{1_p}, \dots, k_{M_p})$. We consider only independent parameters, however, if correlated parameters can be expressed as linear functions then our method can be still applied.

A stochastic system S_p with its stochastic rate constants set to the point $p \in \mathbf{P}$ is represented by a CTMC $C_p = (\mathbb{S}, \mathbf{X}_0, \mathbf{R}_p)$ where transition matrix \mathbf{R}_p is defined as $\mathbf{R}_p(i, j) = \sum_{r \in \text{reac}(\mathbf{X}_i, \mathbf{X}_j)} k_{r_p} \cdot C_{r,i}$. A *set of parametrized CTMCs* induced by the parameter space \mathbf{P} is defined as $\mathbf{C} = \{C_p \mid p \in \mathbf{P}\}$. Henceforward, the states $\mathbf{X}_i \in \mathbb{S}$ will be denoted as s_i .

Uniformization. Uniformization is a standard technique that for a given CTMC $C = (\mathbb{S}, s_0, \mathbf{R})$ computes the transient probability in time t . For an initial state s_0 it returns a vector $\pi^{C, s_0, t}$ such that $\pi^{C, s_0, t}(s') = Pr_{s_0} \{\omega \in \text{Path}^C(s_0) \mid \omega @ t = s'\}$ for all states $s' \in \mathbb{S}$, where Pr_{s_0} is a unique probability measure on all paths ω starting in state s_0 (denoted as $\text{Path}^C(s_0)$) defined, e.g., in [21] and $\omega @ t$ is the state on path ω occupied at time t .

The transient probability in time t is obtained as a sum of expressions giving the state distributions after i discrete reaction steps weighted by the i th Poisson probability $\gamma_{i, q, t} = e^{-q \cdot t} \cdot \frac{(q \cdot t)^i}{i!}$, the probability of i such steps occurring up to t , given the delay is exponentially distributed with *rate* q . Formally, $\pi^{C, s_0, t} = \sum_{i=0}^{\infty} \gamma_{i, q, t} \cdot \pi^{C, s_0, 0} \cdot (\mathbf{Q}^{\text{unif}(C)})^i \approx \sum_{i=L_\varepsilon}^{R_\varepsilon} \gamma_{i, q, t} \cdot \pi^{C, s_0, 0} \cdot (\mathbf{Q}^{\text{unif}(C)})^i$ where $\mathbf{Q}^{\text{unif}(C)}$ is an *uniformized infinitesimal generator matrix* defined as follows: $\mathbf{Q}^{\text{unif}(C)}(s, s') = \frac{\mathbf{R}(s, s')}{q}$, if $s \neq s'$, and $1 - \sum_{s'' \neq s} \frac{\mathbf{R}(s, s'')}{q}$, otherwise, where $q \geq \max\{E^C(s) \mid s \in \mathbb{S}\}$ such that $E^C(s) = \sum_{s' \in \mathbb{S}} \mathbf{R}(s, s')$ is an *exit rate* of the state s in CTMC C . Although the sum is in general infinite, for a given precision ε the upper and lower bounds $L_\varepsilon, R_\varepsilon$ can be estimated by using techniques such as of Fox and Glynn [11] which also allow for efficient computation of Poisson probabilities $\gamma_{i, q, t}$. In order to make the computation feasible the matrix-matrix multiplication is reduced to a vector-matrix multiplication by pre-multiplying, i.e., $\pi^{C, s_0, 0} \cdot (\mathbf{Q}^{\text{unif}(C)})^i = (\pi^{C, s_0, 0} \cdot (\mathbf{Q}^{\text{unif}(C)})^{i-1}) \cdot \mathbf{Q}^{\text{unif}(C)}$.

Property Specification. We consider the *bounded time fragment of CSL with rewards*, see [19] for definition of CSL with rewards. The fragment syntax is defined as follows. A state formula Φ is given as $\Phi ::= \text{true} \mid a \mid \neg\Phi \mid \Phi \wedge \Phi \mid P_{\sim p}[\phi] \mid R_{\sim r}[C^{\leq t}] \mid R_{\sim r}[I^=t]$

where ϕ is a path formula given as $\phi ::= X \Phi \mid \Phi U^I \Phi$, a is an atomic proposition, $\sim \in \{<, \leq, \geq, >\}$, $p \in [0, 1]$ is a probability, $r \in \mathbb{R}_{\geq 0}$ is an expected reward and $I = [a, b]$ is a bounded time interval such that $a, b \in \mathbb{R}_{\geq 0} \wedge a \leq b$. Operators G and F can be derived in the standard way. In order to specify the reward properties, CTMCs are enhanced with reward (cost) structures. Two types of reward structure are used. A *state reward* $\rho(s)$ defines the rate with which a reward is acquired in state $s \in \mathbb{S}$. A reward of $t \cdot \rho(s)$ is acquired if a CTMC remains in state s for t time units. A *transition reward* $\iota(s_i, s_j)$ defines the reward acquired each time the transition (s_i, s_j) occurs.

Let $C = (\mathbb{S}, s_0, \mathbf{R}, L)$ be a labelled CTMC such that L is a labelling function which assigns to each state $s \in \mathbb{S}$ the set $L(s)$ of atomic propositions that are valid in state s .

A state s satisfies $P_{\sim p}[\phi]$ (denoted as $s \models P_{\sim p}[\phi]$) iff $Prob^C(s, \phi) \stackrel{def}{=} Pr_s\{\omega \in Path^C(s) \mid \omega \models \phi\}$ satisfies $\sim p$. A path ω satisfies $X \Phi$ iff $\omega(1) \models \Phi$ where $\omega(1)$ is the second state on ω . A path ω satisfies $\Phi U^I \Psi$ iff $\exists t \in I. (\omega @ t \models \Psi \wedge \forall t' \in [0, t]. (\omega @ t' \models \Phi))$.

Intuitively, a state $s \models R_{\sim p}[C^{\leq t}]$ iff the sum of expected rewards over $Path^C(s)$ *accumulated* until t time units (denoted as $Exp^C(s, X_{C^{\leq t}})$) satisfies $\sim p$. Similarly, a state $s \models R_{\sim p}[I^{=t}]$ iff the sum of expected rewards over all paths $\omega \in Path^C(s)$ at time t (denoted as $Exp^C(s, X_{I^{=t}})$) satisfies $\sim p$. A set $Sat_C(\Phi) = \{s \in \mathbb{S} \mid s \models \Phi\}$ denotes the set of states that satisfy Φ .

The formal semantics of this fragment is defined similarly as the semantics of full CSL and thus we refer the readers to original papers. In the following text all references to CSL address this fragment. Model checking of CSL can be easily reduce to the computation of transient probability, see [3,21] for more details.

3 Parameter Exploration

In this paper we propose an effective method for systematic and fully automatic parameter exploration of a given stochastic system with respect to a specified temporal property and a parameter space. Let \mathbf{C} be a set of parametrized CTMCs describing the dynamics of the stochastic system \mathcal{S} induced by the inspected parameter space \mathbf{P} and a CSL formula Φ expressing the required behaviour. The problem of *parameter exploration* is as follows: for each state $s \in \mathbb{S}$ compute the *landscape function* $\lambda_s^{\Phi, \mathbf{P}} : \mathbf{P} \rightarrow \mathbb{R}_{\geq 0}$ that for each parameter point $p \in \mathbf{P}$ returns the numerical value of the probability or the expected reward for the formula Φ . It means that we consider “quantitative” formulae in the form $\Phi ::= P_{=?}[\phi] \mid R_{=?}[C^{\leq t}] \mid R_{=?}[I^{=t}]$, i.e., the topmost operator of the formula Φ returns a quantitative result, as used, e.g., in PRISM [20]. Note that the formula Φ can contain nested probabilistic and reward operators whose evaluations define discrete sets of states further used in the computation of the resulting numerical value. Therefore, the corresponding landscape function is not in general continuous but only piecewise continuous. Also note that the landscape function is inherently bounded.

To solve the parameter exploration problem we extend *global quantitative model checking* techniques enabling to compute for all states of a CTMC the numerical value of the probability or the expected reward for formula Φ . The most crucial part of the problem is given by the fact that the parameter space \mathbf{P} is continuous and thus the set \mathbf{C} is infinite. Therefore, it is not possible to employ the global quantitative model checking techniques for each CTMC $C_p \in \mathbf{C}$ individually.

Our approach to this problem is based on a new technique which we call *min-max approximation*. The key idea is to approximate the landscape function $\lambda_s^{\Phi, \mathbf{P}}$ using a lower bound $\overline{\min}_s^{\Phi, \mathbf{P}} = \min\{\lambda_s^{\Phi, \mathbf{P}}(p) \mid p \in \mathbf{P}\}$ and an upper bound $\overline{\max}_s^{\Phi, \mathbf{P}} = \max\{\lambda_s^{\Phi, \mathbf{P}}(p) \mid p \in \mathbf{P}\}$. Since the computation of the exact bounds is computationally infeasible, we further approximate these bounds, i.e., we compute approximations $\min_s^{\Phi, \mathbf{P}}$ and $\max_s^{\Phi, \mathbf{P}}$ such that $\min_s^{\Phi, \mathbf{P}} \leq \overline{\min}_s^{\Phi, \mathbf{P}}$ and $\max_s^{\Phi, \mathbf{P}} \geq \overline{\max}_s^{\Phi, \mathbf{P}}$. Although the proposed min-max approximation provides the lower and upper bounds of the landscape function, it introduces an *inaccuracy* with respect to parameter exploration, i.e., such approximation can be insufficient for the inspected parameter space \mathbf{P} and the given formula Φ . Formally, the inaccuracy for a state s is given as the difference $\max_s^{\Phi, \mathbf{P}} - \min_s^{\Phi, \mathbf{P}}$.

A significant advantage of the min-max approximation is that it allows us to iteratively decrease the inaccuracy to a required bound. The key idea is based on *iterative parameter space decomposition* where the parameter space \mathbf{P} is divided into subspaces that are processed independently. The result of such computation is an approximation of the lower bound $\min_s^{\Phi, \mathbf{P}_i}$ and the upper bound $\max_s^{\Phi, \mathbf{P}_i}$ for each subspace \mathbf{P}_i . Such decomposition provides more precise approximation of the landscape function $\lambda_s^{\Phi, \mathbf{P}}$ and enables to reach the required accuracy bound.

In order to effectively compute the min-max approximation for the given formula we design a new method called *parametrized uniformization* allowing to efficiently approximate the transient probabilities for the set \mathbf{C} of parametrized CTMCs. The key idea is to modify *standard uniformization* [14] in such a way that an approximation of the minimal and maximal transient probability with respect to the set \mathbf{C} can be computed. Moreover, the proposed modification preserves the asymptotic time complexity of standard uniformization. Following the model checking method for non-parametrized CTMC presented in [3,21], the result of parametrized uniformization is further used to obtain the min-max approximation of the landscape function $\lambda_s^{\Phi, \mathbf{P}}$.

We are aware that the landscape function could be computed by using standard uniformization to obtain precise values in grid points which could be afterwards interpolated linearly or polynomially. Using adaptive grid refinement such an approach could also provide an arbitrary degree of precision with computation complexity of the same asymptotic class as our method. However, the obtained result would be a general approximation not providing the strict minimal and maximal upper bounds. On the contrary, our min-max approximation guarantees upper and lower estimates without neglecting any singularities caused by possible discontinuities in the landscape function that we consider to be an important feature.

4 Min-Max Approximation

To effectively compute the proposed min-max approximation for an arbitrary nested CSL formula we introduce *the largest and smallest set of states satisfying property Φ* . Let \mathbf{C} be a set of labelled parametrized CTMCs over the parameter space \mathbf{P} such that $\mathbf{C} = \{C_p \mid p \in \mathbf{P}\}$ where each $C_p = (\mathbb{S}, s_0, \mathbf{R}_p, L)$. The maximal set of states satisfying Φ , denoted by $\overline{\text{Sat}}_{\mathbf{C}}^{\top}(\Phi)$, is defined as $\overline{\text{Sat}}_{\mathbf{C}}^{\top}(\Phi) \stackrel{\text{def}}{=} \bigcup_{C_p \in \mathbf{C}} \text{Sat}_{C_p}(\Phi)$. The minimal set of states satisfying Φ , denoted by $\overline{\text{Sat}}_{\mathbf{C}}^{\perp}(\Phi)$, is defined as $\overline{\text{Sat}}_{\mathbf{C}}^{\perp}(\Phi) \stackrel{\text{def}}{=} \bigcap_{C_p \in \mathbf{C}} \text{Sat}_{C_p}(\Phi)$.

Since the set \mathbf{C} is not finite, this definition is not constructive and does not allow to obtain the sets $\overline{Sat}_{\mathbf{C}}^{\top}(\Phi)$ and $\overline{Sat}_{\mathbf{C}}^{\perp}(\Phi)$. Therefore, we define satisfaction relations \models_{\top} and \models^{\perp} that give an alternative characterization of these sets and allow us to effectively compute their approximations.

For any state $s \in \mathbb{S}$ relations $s \models_{\top} \Phi$ and $s \models^{\perp} \Phi$ are defined inductively by:

$$\begin{array}{ll}
s \models_{\top} \text{true} \wedge s \models^{\perp} \text{true}, \text{ for all } s \in \mathbb{S} & s \models_{\top} a \Leftrightarrow s \models^{\perp} a \Leftrightarrow a \in L(s) \\
s \models_{\top} \neg\Phi & \Leftrightarrow s \not\models^{\perp} \Phi \\
s \models_{\top} \Phi \wedge \Psi & \Leftrightarrow s \models_{\top} \Phi \wedge s \models_{\top} \Psi \\
s \models_{\top} P_{\leq p}[\phi] & \Leftrightarrow \overline{Prob}_{\top}^{\mathbf{C}}(s, \phi) \leq p \\
s \models_{\top} P_{\geq p}[\phi] & \Leftrightarrow \overline{Prob}_{\top}^{\mathbf{C}}(s, \phi) \geq p \\
s \models_{\top} R_{\leq p}[I^{\neq t}] & \Leftrightarrow \overline{Exp}_{\top}^{\mathbf{C}}(s, X_{I^{\neq t}}) \leq p \\
s \models_{\top} R_{\geq p}[I^{\neq t}] & \Leftrightarrow \overline{Exp}_{\top}^{\mathbf{C}}(s, X_{I^{\neq t}}) \geq p \\
s \models_{\top} R_{\leq p}[C^{\leq t}] & \Leftrightarrow \overline{Exp}_{\top}^{\mathbf{C}}(s, X_{C^{\leq t}}) \leq p \\
s \models_{\top} R_{\geq p}[C^{\leq t}] & \Leftrightarrow \overline{Exp}_{\top}^{\mathbf{C}}(s, X_{C^{\leq t}}) \geq p \\
s \models^{\perp} \neg\Phi & \Leftrightarrow s \not\models_{\top} \Phi \\
s \models^{\perp} \Phi \wedge \Psi & \Leftrightarrow s \models^{\perp} \Phi \wedge s \models^{\perp} \Psi \\
s \models^{\perp} P_{\leq p}[\phi] & \Leftrightarrow \overline{Prob}_{\perp}^{\mathbf{C}}(s, \phi) \leq p \\
s \models^{\perp} P_{\geq p}[\phi] & \Leftrightarrow \overline{Prob}_{\perp}^{\mathbf{C}}(s, \phi) \geq p \\
s \models^{\perp} R_{\leq p}[I^{\neq t}] & \Leftrightarrow \overline{Exp}_{\perp}^{\mathbf{C}}(s, X_{I^{\neq t}}) \leq p \\
s \models^{\perp} R_{\geq p}[I^{\neq t}] & \Leftrightarrow \overline{Exp}_{\perp}^{\mathbf{C}}(s, X_{I^{\neq t}}) \geq p \\
s \models^{\perp} R_{\leq p}[C^{\leq t}] & \Leftrightarrow \overline{Exp}_{\perp}^{\mathbf{C}}(s, X_{C^{\leq t}}) \leq p \\
s \models^{\perp} R_{\geq p}[C^{\leq t}] & \Leftrightarrow \overline{Exp}_{\perp}^{\mathbf{C}}(s, X_{C^{\leq t}}) \geq p
\end{array}$$

where

$$\begin{aligned}
\overline{Prob}_{\top}^{\mathbf{C}}(s, \phi) &\stackrel{def}{=} \max\{Prob^{C_p}(s, \phi) \mid C_p \in \mathbf{C}\} \\
\overline{Prob}_{\perp}^{\mathbf{C}}(s, \phi) &\stackrel{def}{=} \min\{Prob^{C_p}(s, \phi) \mid C_p \in \mathbf{C}\} \\
\overline{Exp}_{\top}^{\mathbf{C}}(s, X) &\stackrel{def}{=} \max\{Exp^{C_p}(s, X) \mid C_p \in \mathbf{C}\} \text{ for } X \in \{X_{I^{\neq t}}, X_{C^{\leq t}}\} \\
\overline{Exp}_{\perp}^{\mathbf{C}}(s, X) &\stackrel{def}{=} \min\{Exp^{C_p}(s, X) \mid C_p \in \mathbf{C}\} \text{ for } X \in \{X_{I^{\neq t}}, X_{C^{\leq t}}\}
\end{aligned}$$

By structural induction it can be proved that $\forall s \in \mathbb{S} : s \in \overline{Sat}_{\mathbf{C}}^{\top}(\Phi) \Rightarrow s \models_{\top} \Phi$ and $s \models^{\perp} \Phi \Rightarrow s \in \overline{Sat}_{\mathbf{C}}^{\perp}(\Phi)$. This characterization allows us to define an approximation $Sat_{\mathbf{C}}^{\top}(\Phi)$ and $Sat_{\mathbf{C}}^{\perp}(\Phi)$ in the following way. For all $s \in \mathbb{S} : s \in Sat_{\mathbf{C}}^{\top}(\Phi) \stackrel{def}{\Leftrightarrow} s \models_{\top}^* \Phi$ and $s \in Sat_{\mathbf{C}}^{\perp}(\Phi) \stackrel{def}{\Leftrightarrow} s \models^{\perp,*} \Phi$ where the definition of \models_{\top}^* and $\models^{\perp,*}$ differs from the definition of \models_{\top} and \models^{\perp} such that the exact values $\overline{Prob}_{\top}^{\mathbf{C}}(s, \phi)$, $\overline{Prob}_{\perp}^{\mathbf{C}}(s, \phi)$, $\overline{Exp}_{\top}^{\mathbf{C}}(s, X)$ and $\overline{Exp}_{\perp}^{\mathbf{C}}(s, X)$ are replaced by approximate values $Prob_{\top}^{\mathbf{C}}(s, \phi)$, $Prob_{\perp}^{\mathbf{C}}(s, \phi)$, $Exp_{\top}^{\mathbf{C}}(s, X)$ and $Exp_{\perp}^{\mathbf{C}}(s, X)$, respectively, that satisfy the following:

$$\begin{aligned}
Prob_{\top}^{\mathbf{C}}(s, \phi) &\geq \overline{Prob}_{\top}^{\mathbf{C}}(s, \phi) \wedge Prob_{\perp}^{\mathbf{C}}(s, \phi) \leq \overline{Prob}_{\perp}^{\mathbf{C}}(s, \phi) \\
Exp_{\top}^{\mathbf{C}}(s, X) &\geq \overline{Exp}_{\top}^{\mathbf{C}}(s, X) \wedge Exp_{\perp}^{\mathbf{C}}(s, X) \leq \overline{Exp}_{\perp}^{\mathbf{C}}(s, X) \text{ for } X \in \{X_{I^{\neq t}}, X_{C^{\leq t}}\}.
\end{aligned}$$

Since we get that $\forall s \in \mathbb{S} : s \in Sat_{\mathbf{C}}^{\perp}(\Phi) \Rightarrow s \models^{\perp,*} \Phi \Rightarrow s \models^{\perp} \Phi \Rightarrow s \in \overline{Sat}_{\mathbf{C}}^{\perp}(\Phi)$ and also $s \in \overline{Sat}_{\mathbf{C}}^{\top}(\Phi) \Rightarrow s \models_{\top} \Phi \Rightarrow s \models_{\top}^* \Phi \Rightarrow s \in Sat_{\mathbf{C}}^{\top}(\Phi)$, the sets $Sat_{\mathbf{C}}^{\top}(\Phi)$ and $Sat_{\mathbf{C}}^{\perp}(\Phi)$ give us the correct approximations of the sets $\overline{Sat}_{\mathbf{C}}^{\top}(\Phi)$ and $\overline{Sat}_{\mathbf{C}}^{\perp}(\Phi)$, i.e., $\overline{Sat}_{\mathbf{C}}^{\top}(\Phi) \subseteq Sat_{\mathbf{C}}^{\top}(\Phi)$ and $Sat_{\mathbf{C}}^{\perp}(\Phi) \subseteq \overline{Sat}_{\mathbf{C}}^{\perp}(\Phi)$.

In contrast to the exact values their approximations can be efficiently computed using the parametrized uniformization. Therefore, we can also effectively obtain the approximated sets $Sat_{\mathbf{C}}^{\top}(\Phi)$ and $Sat_{\mathbf{C}}^{\perp}(\Phi)$ that are further used in the computation of the

min-max approximation. Formally $\min_s^{\Phi, \mathbf{P}} = \text{Prob}_{\perp}^{\mathbf{C}}(s, \phi)$ and $\max_s^{\Phi, \mathbf{P}} = \text{Prob}_{\top}^{\mathbf{C}}(s, \phi)$ if the topmost operator of the formula Φ is $\mathbf{P}_{=?}[\phi]$. Similarly, $\min_s^{\Phi, \mathbf{P}} = \text{Exp}_{\perp}^{\mathbf{C}}(s, \mathbf{X})$ and $\max_s^{\Phi, \mathbf{P}} = \text{Exp}_{\top}^{\mathbf{C}}(s, \mathbf{X})$ for $\mathbf{X} = \mathbf{X}_{\mathbf{C} \leq t}$ and $\mathbf{X} = \mathbf{X}_{\mathbf{I} = t}$ if the topmost operator of the formula Φ is $\mathbf{R}_{=?}[\mathbf{C} \leq t]$ and $\mathbf{R}_{=?}[\mathbf{I} = t]$, respectively.

5 Parametrized Uniformization

The most important step of the proposed min-max approximation is for each state s to compute the values $\text{Prob}_{\top}^{\mathbf{C}}(s, \phi)$, $\text{Prob}_{\perp}^{\mathbf{C}}(s, \phi)$, $\text{Exp}_{\top}^{\mathbf{C}}(s, \mathbf{X})$ and $\text{Exp}_{\perp}^{\mathbf{C}}(s, \mathbf{X})$ where \mathbf{C} is an infinite set of parametrized CTMCs, ϕ is an arbitrary path formula and $\mathbf{X} \in \{\mathbf{X}_{\mathbf{C} \leq t}, \mathbf{X}_{\mathbf{I} = t}\}$. In order to efficiently obtain these values we employ parametrized uniformization. It is a technique that for the given set \mathbf{C} , state $s \in \mathbb{S}$ and time $t \in \mathbb{R}_{\geq 0}$ computes vectors $\pi_{\top}^{\mathbf{C}, s, t}$ and $\pi_{\perp}^{\mathbf{C}, s, t}$ such that for each state $s' \in \mathbb{S}$ the following holds:

$$\pi_{\top}^{\mathbf{C}, s, t}(s') \geq \max\{\pi^{C_p, s, t}(s') \mid C_p \in \mathbf{C}\} \wedge \pi_{\perp}^{\mathbf{C}, s, t}(s') \leq \min\{\pi^{C_p, s, t}(s') \mid C_p \in \mathbf{C}\} \quad (1)$$

The key idea of parametrized uniformization is to modify standard uniformization in such a way that for each state s' and in each iteration i of the computation we locally minimize (maximize) the value $\pi^{\mathbf{C}, s, t}(s')$ with respect to each $C_p \in \mathbf{C}$. It means that in the i th iteration of the computation for a state s' we consider only the minimal (maximal) values of the relevant states in the iteration $i - 1$, i.e., the states that affect the value of state s' . We show that the local minimum and maximum can be efficiently computed and that it gives us values satisfying Equation 1.

Let $\mathbf{Q}^{\text{unif}(\mathbf{C})}$ be an uniformized infinitesimal generator matrix for a set \mathbf{C} of parametrized CTMCs defined as follows:

$$\mathbf{Q}^{\text{unif}(\mathbf{C})}(i, j) = \begin{cases} \sum_{r \in \text{reac}(\mathbf{X}_i, \mathbf{X}_j)} k_r \cdot \frac{C_{ri}}{q_{\max}} & \text{if } i \neq j \\ 1 - \sum_{l \neq i} \sum_{r \in \text{reac}(\mathbf{X}_i, \mathbf{X}_l)} k_r \cdot \frac{C_{ri}}{q_{\max}} & \text{otherwise.} \end{cases} \quad (2)$$

where $q_{\max} \geq E_{\max} = \max\{E^{C_p}(s) \mid C_p \in \mathbf{C}, s \in \mathbb{S}\}$ and k_r is a variable from $[k_r^{\perp}, k_r^{\top}]$.

For sake of simplicity, we present only the method allowing to efficiently compute the vector $\pi_{\top}^{\mathbf{C}, s, t}$, since the computation of $\pi_{\perp}^{\mathbf{C}, s, t}$ is symmetric. We start with the trivial observation that vectors $\pi^{C_p, s, 0}$ (initial probability distributions, e.g., $\pi^{C_p, s, 0}(s') = 1$, if $s = s'$, and 0, otherwise) are equal for all $C_p \in \mathbf{C}$. Therefore $\pi^{\mathbf{C}, s, 0} = \pi_{\top}^{\mathbf{C}, s, 0} = \pi^{C_p, s, 0}$ for all $C_p \in \mathbf{C}$. In order to present parametrized uniformization, we introduce an operator \odot_{\top} such that for each $s' \in \mathbb{S}$ the following holds:

$$\left(\pi_{\top}^{\mathbf{C}, s, 0} \odot_{\top} \left(\mathbf{Q}^{\text{unif}(\mathbf{C})} \right)^i \right) (s') \geq \max \left\{ \left(\pi^{C_p, s, 0} \cdot \left(\mathbf{Q}^{\text{unif}(C_p)} \right)^i \right) (s') \mid C_p \in \mathbf{C} \right\}.$$

Moreover, we further require that vectors from the previous iteration can be used, in particular, $\pi_{\top}^{\mathbf{C}, s, 0} \odot_{\top} \left(\mathbf{Q}^{\text{unif}(\mathbf{C})} \right)^i = \left(\pi_{\top}^{\mathbf{C}, s, 0} \odot_{\top} \left(\mathbf{Q}^{\text{unif}(\mathbf{C})} \right)^{i-1} \right) \odot_{\top} \mathbf{Q}^{\text{unif}(\mathbf{C})}$. The operator \odot_{\top} returns a vector $\pi'_{\top} \in \mathbb{R}_{\geq 0}^{|\mathbb{S}|}$ containing for each state $s_i \in \mathbb{S}$ the maximal possible probability after a single discrete step of a DTMC obtained by uniformization of any CTMC

C_p , i.e., $\left(\pi \odot_{\top} \mathbf{Q}^{\text{unif}(\mathbf{C})}\right)(s) \stackrel{\text{def}}{=} \max \left\{ \left(\pi \cdot \mathbf{Q}^{\text{unif}(C_p)}\right)(s) \mid p \in \mathbf{P} \right\} = \pi'_{\top}(s)$ where π is a general vector. Since $\sum \pi'_{\top}(i) \geq 1$, the vector π'_{\top} is no longer a state distribution.

To show how the operator \odot_{\top} is computed let $\sigma(s_i)$ be an algebraic expression defined as the part of the vector-matrix multiplication for state s_i . For each $s_i \in \mathbb{S}$ we get $\sigma(s_i) = \left(\pi \cdot \mathbf{Q}^{\text{unif}(\mathbf{C})}\right)(s_i) = \sum_{j=0}^{|\mathbb{S}|-1} \pi(j) \cdot \mathbf{Q}^{\text{unif}(\mathbf{C})}(j, i)$. Rewriting $\sigma(s_i)$ by Equation 2 and using the sets $\text{pred}(s_i)$ and $\text{succ}(s_i)$ we obtain the following:

$$\sigma(s_i) = \sum_{(j,r) \in \text{pred}(s_i)} \pi(j) \cdot k_r \cdot \frac{C_{r,j}}{q_{\max}} + \pi(i) \left(1 - \sum_{(j,r) \in \text{succ}(s_i)} k_r \cdot \frac{C_{r,i}}{q_{\max}} \right) \quad (3)$$

The first summand in Equation 3, indexed over predecessors of s_i , corresponds to the probability mass inflowing into state s_i through all reactions. The second summand corresponds to the portion of probability mass remaining in s_i from the previous iteration.

The operator \odot_{\top} locally maximizes expression $\sigma(s)$ for all $s \in \mathbb{S}$ with respect to \mathbf{P} , i.e., $\left(\pi \odot_{\top} \mathbf{Q}^{\text{unif}(\mathbf{C})}\right)(s) = \max \{ \sigma_p(s) \mid p \in \mathbf{P} \}$ where $\sigma_p(s)$ is the evaluation of $\sigma(s)$ in the parameter point $p = (k_{1,p}, \dots, k_{M,p})$. First, we show that to compute expression $\pi_{\top}^{\mathbf{C},s,0} \odot_{\top} (\mathbf{Q}^{\text{unif}(\mathbf{C})})^i$ it is sufficient to consider only maximal values from the previous iteration, i.e., vector $\pi_{\top}^{\mathbf{C},s,0} \odot_{\top} (\mathbf{Q}^{\text{unif}(\mathbf{C})})^{i-1}$. Note that $\forall s_i \in \mathbb{S}. \pi(i) \geq 0$ and $\forall (j,r) \in \text{pred}(s_i) \cup \text{succ}(s_i). k_r \geq 0 \wedge C_{r,j} \geq 0$. Moreover, since $q_{\max} \geq E_{\max} \geq 0$, we get that $(1 - \sum_{(j,r) \in \text{succ}(s_i)} k_r \cdot \frac{C_{r,i}}{q_{\max}}) \geq (1 - \frac{E_{\max}}{q_{\max}}) \geq 0$. Now, we can see from Equation 3 that in order to maximize $\sigma(s_i)$ maximal values of $\pi(i)$ for each $0 \leq i < |\mathbb{S}|$ have to be taken.

Second, we show how to determine $p = \{k_1, \dots, k_M\} \in \mathbf{P}$ such that $\sigma(s_i)$ is evaluated as a maximum. Equation 3 can be rewritten in the following way:

$$\sigma(s_i) = \sum_{(j,r) \in \text{in}} k_r \cdot \frac{\pi(j) \cdot C_{r,j}}{q_{\max}} + \sum_{(j,r) \in \text{inout}} k_r \cdot \frac{\pi(j) \cdot C_{r,j} - \pi(i) \cdot C_{r,i}}{q_{\max}} - \sum_{(j,r) \in \text{out}} k_r \cdot \frac{\pi(i) \cdot C_{r,i}}{q_{\max}}$$

where $\text{in} = \text{pred}(s_i) \setminus \text{succ}(s_i)$, $\text{inout} = \text{pred}(s_i) \cap \text{succ}(s_i)$ and $\text{out} = \text{succ}(s_i) \setminus \text{pred}(s_i)$. The three sums range over disjoint sets of reactions. The first sum represents all incoming reactions that do not have an outgoing counterpart, for these $k_r = k_r^{\top}$, since they only increase $\sigma(s_i)$. The second sum represents reactions flowing into s_i as well as flowing out of s_i . In this case, expression $\pi(j) \cdot C_{r,j} - \pi(i) \cdot C_{r,i}$ has to be evaluated. If it is positive, $k_r = k_r^{\top}$, otherwise, $k_r = k_r^{\perp}$. The last sum represents only reactions flowing out of s_i and hence $k_r = k_r^{\perp}$. The operator \odot_{\top} is now computed as $\left(\pi \odot_{\top} \mathbf{Q}^{\text{unif}(\mathbf{C})}\right)(s_i) = \sigma(s_i)$ where each k_r inside $\sigma(s_i)$ is chosen according to the aforementioned rules.

The computation of $\pi_{\perp}^{\mathbf{C},s,t}$ is symmetric to the case of $\pi_{\top}^{\mathbf{C},s,t}$. It means that we define the operator \odot_{\perp} which locally minimize expression $\sigma(s)$ for all $s \in \mathbb{S}$ with respect to \mathbf{P} . In order to minimize $\sigma(s_i)$ it is sufficient to inverse the aforementioned rules.

Vectors $\pi_{\top}^{\mathbf{C},s,t}$ and $\pi_{\perp}^{\mathbf{C},s,t}$ are now computed similarly as in the case of standard uniformization, i.e., $\pi_{\star}^{\mathbf{C},s,t} = \sum_{i=L_{\epsilon}}^{R_{\epsilon}} \gamma_{i,q,t} \cdot \pi^{\mathbf{C},s,0} \odot_{\star} (\mathbf{Q}^{\text{unif}(\mathbf{C})})^i$ where $\star \in \{\perp, \top\}$. To obtain the required values $\text{Prob}_{\top}^{\mathbf{C}}(s, \phi)$, $\text{Prob}_{\perp}^{\mathbf{C}}(s, \phi)$, $\text{Exp}_{\top}^{\mathbf{C}}(s, X)$ and $\text{Exp}_{\perp}^{\mathbf{C}}(s, X)$, we employ the standard model checking technique [3,21] where transient probability $\pi^{\mathbf{C},s,t}$ for a non-parametrized CTMC C is replaced by vectors $\pi_{\top}^{\mathbf{C},s,t}$ and $\pi_{\perp}^{\mathbf{C},s,t}$.

Compared to standard uniformization, only a constant amount of additional work has to be performed in order to determine parameter values. Therefore, asymptotic complexity of parametrized uniformization remains the same as standard uniformization.

6 Parameter Space Decomposition

Before we describe parameter space decomposition – a method allowing to reduce the inaccuracy of the proposed min-max approximation – we briefly discuss the key characteristics of parametrized uniformization. The most important fact is that parametrized uniformization for the set \mathbf{C} in general does not correspond to standard uniformization for any CTMC $C_p \in \mathbf{C}$. The reason is that we consider a behaviour of a parametrized CTMC that has no equivalent counterpart in any particular C_p . First, the parameter k_r in Equation 3 is determined individually for each state. Therefore, in a single iteration $k_r = k_r^\top$ for one state and $k_r = k_r^\perp$ for another state. Second, the parameter is determined individually for each iteration and thus for a state s_i the parameter k_r can be chosen differently in individual iterations.

Inaccuracy of the proposed min-max approximation related to the computation of parametrized uniformization, called *unification error*, is given as $(\max_s^{\Phi, \mathbf{P}} - \overline{\max}_s^{\Phi, \mathbf{P}}) + (\overline{\min}_s^{\Phi, \mathbf{P}} - \min_s^{\Phi, \mathbf{P}})$. Apart from the unification error our approach introduces an inaccuracy related to approximation of the landscape function $\lambda_s^{\Phi, \mathbf{P}_i}$, called *approximation error*, given as $\overline{\max}_s^{\Phi, \mathbf{P}} - \overline{\min}_s^{\Phi, \mathbf{P}}$. Finally, the *overall error* of the min-max approximation, denoted as $\text{Err}_s^{\Phi, \mathbf{P}}$, is defined as a sum of both errors, i.e., $\text{Err}_s^{\Phi, \mathbf{P}} = \max_s^{\Phi, \mathbf{P}} - \min_s^{\Phi, \mathbf{P}}$. Fig. 1 illustrates both types of errors. The approximation error is depicted as blue rectangles and the unification error is depicted as the red rectangles.

We are not able to effectively distinguish the proportion of the approximation error and the unification error nor to reduce the unification error as such. Therefore, we design a method based on the parameter space decomposition that allows us to effectively reduce the overall error of the min-max approximation to a user specified *absolute error bound*, denoted as ERR .

In order to ensure that the min-max approximation meets the given absolute error bound ERR , we iteratively decompose the parameter space \mathbf{P} into finitely many subspaces such that $\mathbf{P} = \mathbf{P}_1 \cup \dots \cup \mathbf{P}_n$ and each partial result satisfies the overall error bound, i.e., $\forall s \in \mathbb{S} : \max_s^{\Phi, \mathbf{P}_i} - \min_s^{\Phi, \mathbf{P}_i} \leq \text{ERR}$. Therefore, the overall error for each state $s \in \mathbb{S}$ equals to $\text{Err}_s^{\Phi, \mathbf{P}} = \sum_{i=1}^n \frac{|\mathbf{P}_i|}{|\mathbf{P}|} (\max_s^{\Phi, \mathbf{P}_i} - \min_s^{\Phi, \mathbf{P}_i}) \leq \sum_{i=1}^n \frac{|\mathbf{P}_i|}{|\mathbf{P}|} \text{ERR} = \text{ERR}$. Fig. 1 illustrates such a decomposition and demonstrates convergence of $\text{Err}_s^{\Phi, \mathbf{P}_i}$ to 0 provided that the function $\lambda_s^{\Phi, \mathbf{P}_i}$ is continuous.

For sake of simplicity, we present parametric decomposition on the computation of $\pi_{\top}^{\mathbf{C}, s, t}$ since it can be easily extended to the computation of $\text{Prob}_{\top}^{\mathbf{C}}(s, \phi)$, $\text{Prob}_{\perp}^{\mathbf{C}}(s, \phi)$, $\text{Exp}_{\top}^{\mathbf{C}}(s, \mathbf{X})$ and $\text{Exp}_{\perp}^{\mathbf{C}}(s, \mathbf{X})$. If during the computation in an iteration i for a state $s' \in \mathbb{S}$ holds that $(\pi_{\top}^{\mathbf{C}, s, 0} \odot_{\top} (\mathbf{Q}^{\text{unif}(\mathbf{C})})^i)(s') - (\pi_{\perp}^{\mathbf{C}, s, 0} \odot_{\perp} (\mathbf{Q}^{\text{unif}(\mathbf{C})})^i)(s') > \text{ERR}$ we cancel the current computation and decompose the parameter space \mathbf{P} to n subspaces such that $\mathbf{P} = \mathbf{P}_1 \cup \dots \cup \mathbf{P}_n$. Each subspace \mathbf{P}_j defines a new set of CTMCs $\mathbf{C}_j = \{C_j \mid j \in \mathbf{P}_j\}$ that is independently processed in a new computation branch. Note that we could

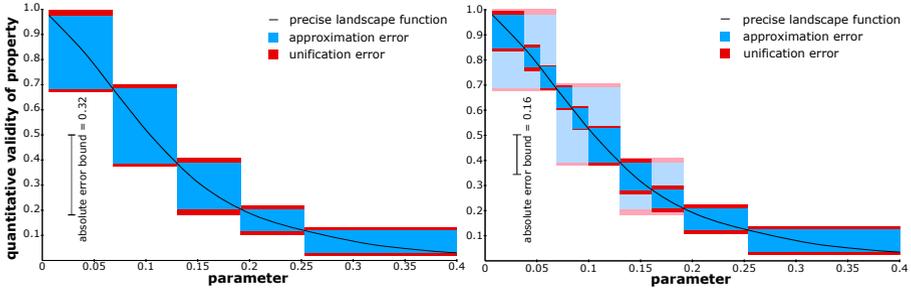


Fig. 1. Illustration of the min-max approximation computation of the landscape function $\lambda_s^{\Phi, \mathbf{P}}$ for an initial state s , property Φ and parameter space $\mathbf{P} = [0, 0.4]$. Left graph shows the decomposition of \mathbf{P} into 5 subspaces for absolute error bound $\text{ERR} = 0.32$. Right graph shows a more refined decomposition for $\text{ERR} = 0.16$ resulting in 10 subspaces. This decomposition reduces both types of errors in each refined subspaces. The exact shape of $\lambda_s^{\Phi, \mathbf{P}}$ is visualized as the black curve.

reuse the previous computation and continue from the iteration $i - 1$. However, the most significant part of the error is usually cumulated during the the previous iterations and thus the decomposition would have only a negligible impact on error reduction.

A *minimal decomposition with respect to the parameter space \mathbf{P}* defines a minimal number of subspaces m such that $\mathbf{P} = \mathbf{P}_1 \cup \dots \cup \mathbf{P}_m$ and for each subspace \mathbf{P}_j where $1 \leq j \leq m$ holds that $\text{Err}_s^{\Phi, \mathbf{P}_j} \leq \text{ERR}$. Note that the existence of such decomposition is guaranteed only if the landscape function $\lambda_s^{\Phi, \mathbf{P}}$ is continuous. If the landscape function is continuous there can exist more than one minimal decomposition. However, it can not be straightforwardly found. To overcome this problem we have considered and implemented several heuristics allowing to iteratively compute a decomposition satisfying the following: (1) it ensures the required error bound whenever $\lambda_s^{\Phi, \mathbf{P}}$ is continuous, (2) it guarantees the refinement termination in the situation where $\lambda_s^{\Phi, \mathbf{P}}$ is not continuous and the discontinuity causes that ERR can not be achieved. To ensure the termination an additional parameter has to be introduced as a lower bound on the subspace size. Hence this parameter provides a supplementary termination criterion.

7 Case Studies

We implemented our method on top of the tool PRISM 4.0 [20]. We run all experiments on a Linux workstation with an AMD PhenomTM II X4 940 Processor @ 3GHz, 8 GB DDR2 @ 1066 MHz RAM. We used PRISM version 4.0.3. running with sparse engine, since this engine is typically faster than its symbolic counterparts due to efficient matrix vector multiplication.

Schloegl's Model. We use Schloegl's model [24] to demonstrate the practicability of our method for parameter exploration with respect to basic transient analysis. It is the simplest biochemical reaction model for which stochasticity is crucial due to bi-stability

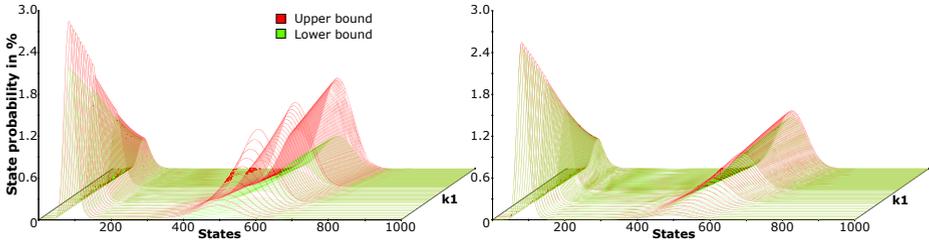


Fig. 2. Species X distribution at 20 time units for $k_1 \in [0.029, 0.031]$ (in s^{-1}). The two presented cases differ in absolute error bound: (left) $ERR = 0.01$, (right) $ERR = 0.001$.

– existence of two different steady states to which the species population can (non-deterministically) converge. The model is defined by the following reactions [8]: $2X \xrightarrow{k_1} 3X$, $3X \xrightarrow{k_2} 2X$, $\emptyset \xrightarrow{k_3} X$, $X \xrightarrow{k_4} \emptyset$; $k_1 = 0.03s^{-1}$, $k_2 = 10^{-4}s^{-1}$, $k_3 = 200s^{-1}$, $k_4 = 3.5s^{-1}$. Deterministic formulation of the model by means of ordinary differential equations (ODE) predicts for $k_1 \in [0.0285, 0.035]$ two steady states to which the population converges in the horizon of 20 time units. We will focus on the range $k_1 \in [0.029, 0.031]$. Under the deterministic setting, from any initial state the dynamics evolves to a single steady state. In the noisy setting [26], the population of molecules distributes around both steady states (in short time perspective, here 20 time units). In long time perspective, the population oscillates around both steady states.

We focus on the short time-scale – to analyze the population of X at time 20 starting from the initial state where the number of X is 250. According to the respective ODE model, the population always converges to an asymptotic steady state $X_{st} \leq 1000$. Considering this as an assumption it allows us to bound the state space. The corresponding CTMC has 1001 states and 2000 transitions. The goal of the analysis is to explore how the observed distribution is affected when perturbing k_1 in the range $[0.029, 0.031]$. By executing our method for the absolute error bound $ERR = 0.01$ we got the result visualized in Fig. 2 (left). It can be directly seen that for each parameter point there is a non-zero probability that some individuals reside near the higher steady state while some reside near the lower steady state at time 20. The jumps that are mostly observable in distributions around the higher steady state are caused by the approximation error. Computation with a one order lower error gives a smooth result, see Fig. 2 (right).

The computation required $7.36 \cdot 10^5$ iterations of the parametrized uniformization. The parameter decomposition resulted in 76 subspaces for $ERR = 0.01$ and 639 subspaces for $ERR = 0.001$. The overall computation took 2 and 16.5 hours, respectively.

Gene Regulation of Mammalian Cell Cycle. We have applied the min-max approximation to the gene regulation model published in [25], the regulatory network is shown in Fig. 3a. The model explains regulation of a transition between early phases of the mammalian cell cycle. In particular, it targets the transition from the control G_1 -phase to S -phase (the synthesis phase). G_1 -phase makes an important checkpoint controlled by a *bistable regulatory circuit* based on an interplay of the retinoblastoma protein pRB , denoted by A (the so-called tumour suppressor, HumanCyc:HS06650) and the

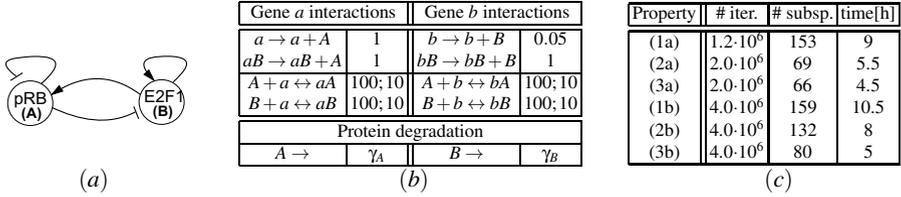


Fig. 3. (a) Two-gene regulatory circuit controlling G_1/S transition in mammalian cell cycle. (b) Stochastic mass action model of the G_1/S regulatory circuit – a, b represent genes, aA, aB, bA, bB represent transcription factor–gene promoter complexes (c) Computation results.

retinoblastoma-binding transcription factor E_2F_1 , denoted by B (a central regulator of a large set of human genes, HumanCyc:HS02261). In high concentration levels, the E_2F_1 protein activates the G_1/S transition mechanism. On the other hand, a low concentration of E_2F_1 prevents committing to S -phase.

Positive autoregulation of B causes bi-stability of its concentration depending on the parameters. Especially, of specific interest is the degradation rate of A , γ_A . In [25] it is shown that for increasing γ_A the low stable mode of B switches to the high stable mode. When mitogenic stimulation increases under conditions of active growth, rapid phosphorylation of A starts and makes the degradation of unphosphorylated A stronger (the degradation rate γ_A increases). This causes B to lock in the high stable mode implying the cell cycle commits to S -phase. Since mitogenic stimulation influences the degradation rate of A , our goal is to study the population distribution around the low and high steady state and to explore the effect of γ_A by means of the landscape function.

We have translated the original ODE model into the framework of stochastic mass action kinetics [12]. The resulting reactions are shown in Fig. 3b. Since the detailed knowledge of elementary chemical reactions occurring in the process of transcription and translation is incomplete, we use the simplified form as suggested in [10]. In the minimalistic setting, the reformulation requires addition of rate parameters describing the transcription factor–gene promoter interaction while neglecting cooperativeness of transcription factors activity. Our parametrization is based on time-scale orders known for the individual processes [27] (parameters considered in s^{-1}). Moreover, we assume the numbers of A and B are bounded by 10 molecules. Upper bounds for A and B are set with respect to behaviour of an ensemble of stochastic simulations. We consider minimal population number distinguishing the two stable modes. All other species are bounded by the initial number of DNA molecules (genes a and b) which is conserved and set to 1. The corresponding CTMC has 1078 states and 5919 transitions.

We consider three hypotheses: (1) stabilization in the low mode where $B < 3$, (2) stabilization in the high mode where $B > 5$, (3) stabilization in the high mode where $B > 7$ ((3) is more focused than (2)). All the hypotheses are expressed within time horizon 1000 seconds reflecting the time scale of gene regulation response. We employ two alternative CSL formulations to express each of the three hypothesis. According to [25], we consider the parameter space $\gamma_A \in [0.005, 0.5]$.

First, we express the property of being inside the given bound during the time interval $I = [500, 1000]$ using globally operator: (1a) $P_{\sim?}[G^I(B < 3)]$, (2a) $P_{\sim?}[G^I(B > 5)]$ and (3a) $P_{\sim?}[G^I(B > 7)]$. The interval starts from 500 seconds in order to bridge the

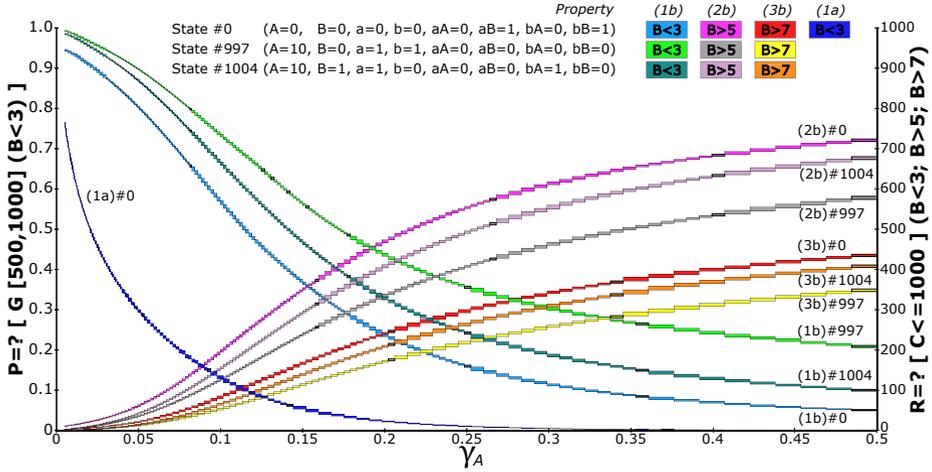


Fig. 4. Landscape functions of properties (1a, 1b, 2b, 3b) for $\gamma_A \in [0.005, 0.5]$ (in s^{-1}) and initial states #0, #997 and #1004. The left Y-axis scale corresponds to (1a), the right to (1b, 2b, 3b).

initial fluctuation region and let the system stabilize. Since the stochastic noise causes molecules to repeatedly escape the requested bound, the resulting probability is significantly lower than expected. Namely, in cases (2a) and (3a) the resulting probability is close to 0 for the whole parameter space. Moreover, the selection of an initial state has only a negligible impact on the result. Therefore, in Fig. 4 only the resulting probability for case (1a) and a single selected initial state is visualized.

Second, we use a cumulative reward property to capture the fraction of the time the system has the required number of molecules within the time interval $[0, 1000]$: (1b) $R_{\sim?}[C^{\leq t}](B < 3)$, (2b) $R_{\sim?}[C^{\leq t}](B > 5)$, (3b) $R_{\sim?}[C^{\leq t}](B > 7)$ where $t = 1000$ and $R_{\sim?}[C^{\leq t}](B \sim X)$ denotes that state reward ρ is defined such that $\forall s \in \mathbb{S}, \rho(s) = 1$ iff $B \sim X$ in s . The result is visualized for three selected initial states in Fig. 4.

Fig. 4 also illustrates inaccuracy of our approach with respect to the absolute error bound $ERR = 0.01$ by means of small rectangles depicting approximations of the resulting probabilities and expected rewards. The analyses predict that the distribution of the low steady mode interferes with the distribution of the high steady mode. It confirms bi-stability predicted in [25] but in contrast to ODE analysis our method shows how the population of cells distributes around the two stable states. Results of computations including the number of iterations performed during parametrized uniformization, numbers of resulting subspaces and execution times in hours, are presented in Fig. 3c.

Finally, to see how degradation rates of A and B cooperate in affecting property (3b), we explore two-dimensional parameter space $(\gamma_A, \gamma_B) \in [0.005, 0.1] \times [0.05, 0.1]$. The computation also required $4.0 \cdot 10^6$ iterations of the parametrized uniformization, the parameter decomposition resulted in 143 subspaces for $ERR = 0.1$ and the overall execution took 14 hours. Fig. 5 illustrates the computed upper bound of the landscape function for initial state #0 and the absolute error. The result predicts antagonistic relation between the degradation rates which is in agreement with the ODE model [25].

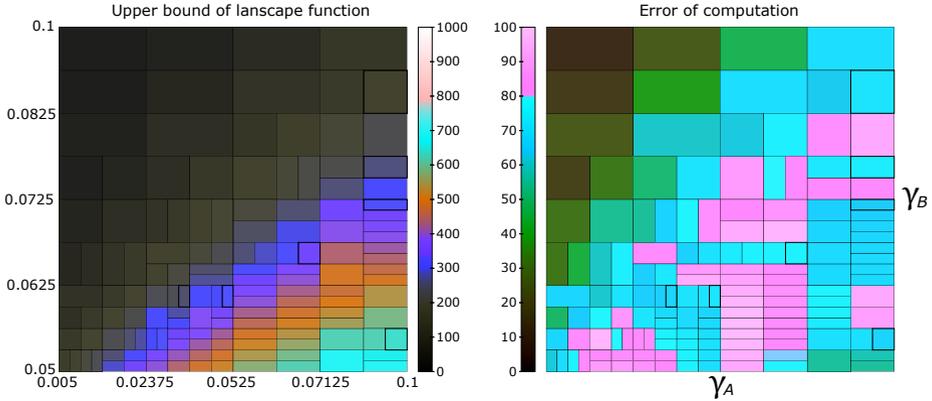


Fig. 5. Landscape function for property (3b), initial state #0 ($A = 0, B = 0, a = 0, b = 0, aA = 0, aB = 1, bA = 0, bB = 1$) and two-dimensional parameter space $(\gamma_A, \gamma_B) \in [0.005, 0.1] \times [0.05, 0.1]$ (represented in s^{-1} by X and Y axes, respectively). On the left, the upper bound of the landscape function is illustrated. On the right, the absolute error given as difference between computed upper and lower bounds is depicted. In both cases the color scale is used.

8 Conclusions

We have introduced the parameter exploration problem for stochastic biochemical systems as the computation of a landscape function for a given temporal logic formula. The key idea of our approach is to approximate the lower and upper bounds of the landscape function. To obtain such approximation for an arbitrary nested CSL formula, we compute the largest and smallest set of states satisfying the formula using parametrized uniformization. This allows to approximate the minimal and maximal transient probability with respect to the parameter space. In order to reach a required error bound of the proposed approximation, we iteratively decompose the parameter space and compute the approximation for each subspace individually. We have demonstrated our approach to the parameter exploration problem on two biologically motivated case studies.

The experiments show that our method can be extremely time demanding and thus in our future work we will focus on its acceleration. We plan to apply techniques allowing to accelerate the underlying transient analysis [9,16] and more efficient heuristics for the parameter space decomposition. Moreover, our method can be easily parallelized and thus a significant acceleration can be obtained.

References

1. Andreychenko, A., Mikeev, L., Spieler, D., Wolf, V.: Parameter Identification for Markov Models of Biochemical Reactions. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 83–98. Springer, Heidelberg (2011)
2. Aziz, A., Sanwal, K., Singhal, V., Brayton, R.: Verifying Continuous Time Markov Chains. In: Alur, R., Henzinger, T.A. (eds.) CAV 1996. LNCS, vol. 1102, pp. 269–276. Springer, Heidelberg (1996)

3. Baier, C., Haverkort, B., Hermanns, H., Katoen, J.P.: Model Checking Continuous-Time Markov Chains by Transient Analysis. In: Emerson, E.A., Sistla, A.P. (eds.) CAV 2000. LNCS, vol. 1855, pp. 358–372. Springer, Heidelberg (2000)
4. Ballarini, P., Forlin, M., Mazza, T., Prandi, D.: Efficient Parallel Statistical Model Checking of Biochemical Networks. In: PDMC 2009. EPTCS, vol. 14, pp. 47–61 (2009)
5. Barbuti, R., Levi, F., Milazzo, P., Scatena, G.: Probabilistic Model Checking of Biological Systems with Uncertain Kinetic Rates. *Theor. Comput. Sci.* 419, 2–16 (2012)
6. Bernardini, F., Biggs, C., Derrick, J., Gheorghe, M., Niranjani, M., Sanguinetti, G.: Parameter Estimation and Model Checking in a Model of Prokaryotic Autoregulation. Tech. rep., University of Sheffield (2007)
7. Daigle, B., Roh, M., Petzold, L., Niemi, J.: Accelerated Maximum Likelihood Parameter Estimation for Stochastic Biochemical Systems. *BMC Bioinformatics* 13(1), 68–71 (2012)
8. Degasperis, A., Gilmore, S.: Sensitivity Analysis of Stochastic Models of Bistable Biochemical Reactions. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) SFM 2008. LNCS, vol. 5016, pp. 1–20. Springer, Heidelberg (2008)
9. Didier, F., Henzinger, T.A., Mateescu, M., Wolf, V.: Fast Adaptive Uniformization of the Chemical Master Equation. In: HIBI 2009, pp. 118–127. IEEE Computer Society (2009)
10. El Samad, H., Khammash, M., Petzold, L., Gillespie, D.: Stochastic Modelling of Gene Regulatory Networks. *Int. J. of Robust and Nonlinear Control* 15(15), 691–711 (2005)
11. Fox, B.L., Glynn, P.W.: Computing Poisson Probabilities. *CACM* 31(4), 440–445 (1988)
12. Gillespie, D.T.: Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry* 81(25), 2340–2381 (1977)
13. Golightly, A., Wilkinson, D.J.: Bayesian Parameter Inference for Stochastic Biochemical Network Models Using Particle Markov Chain Monte Carlo. *Interface Focus* 1(6), 807–820 (2011)
14. Grassmann, W.: Transient Solutions in Markovian Queueing Systems. *Computers & Operations Research* 4(1), 47–53 (1977)
15. Hahn, E.M., Han, T., Zhang, L.: Synthesis for PCTL in Parametric Markov Decision Processes. In: NASA Formal Methods, pp. 146–161 (2011)
16. Henzinger, T.A., Mateescu, M., Wolf, V.: Sliding Window Abstraction for Infinite Markov Chains. In: Bouajjani, A., Maler, O. (eds.) CAV 2009. LNCS, vol. 5643, pp. 337–352. Springer, Heidelberg (2009)
17. Jha, S.K., Clarke, E.M., Langmead, C.J., Legay, A., Platzer, A., Zuliani, P.: A Bayesian Approach to Model Checking Biological Systems. In: Degano, P., Gorrieri, R. (eds.) CMSB 2009. LNCS, vol. 5688, pp. 218–234. Springer, Heidelberg (2009)
18. Koh, C.H., Palaniappan, S., Thiagarajan, P., Wong, L.: Improved Statistical Model Checking Methods for Pathway Analysis. *BMC Bioinformatics* 13(suppl. 17), S15 (2012)
19. Kwiatkowska, M., Norman, G., Pacheco, A.: Model Checking Expected Time and Expected Reward Formulae with Random Time Bounds. *Compu. Math. Appl.* 51(2), 305–316 (2006)
20. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: Verification of Probabilistic Real-time Systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 585–591. Springer, Heidelberg (2011)
21. Kwiatkowska, M., Norman, G., Parker, D.: Stochastic Model Checking. In: Bernardo, M., Hillston, J. (eds.) SFM 2007. LNCS, vol. 4486, pp. 220–270. Springer, Heidelberg (2007)
22. Mikeev, L., Neuhäuser, M., Spieler, D., Wolf, V.: On-the-fly Verification and Optimization of DTA-properties for Large Markov Chains. *Form. Method. Syst. Des.*, 1–25 (2012)
23. Reinker, S., Altman, R., Timmer, J.: Parameter Estimation in Stochastic Biochemical Reactions. *IEEE Proc. Syst. Biol.* 153(4), 168–178 (2006)

24. Schlögl, F.: Chemical Reaction Models for Non-Equilibrium Phase Transitions. *Zeitschrift für Physik* 253, 147–161 (1972)
25. Swat, M., Kel, A., Herzog, H.: Bifurcation Analysis of the Regulatory Modules of the Mammalian G1/S transition. *Bioinformatics* 20(10), 1506–1511 (2004)
26. Vellela, M., Qian, H.: Stochastic Dynamics and Non-Equilibrium Thermodynamics of a Bistable Chemical System: the Schlögl Model Revisited. *Journal of The Royal Society Interface* 6(39), 925–940 (2009)
27. Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M.K., Magnasco, M., Darnell, J.E.: Decay Rates of Human mRNAs: Correlation With Functional Characteristics and Sequence Attributes. *Genome Research* 13(8), 1863–1872 (2003)