Bounding Mean First Passage Times in Population Continuous-Time Markov Chains

Michael Backenköhler[1], Luca Bortolussi[2,1], Verena Wolf[1]

[1]Saarland University, Germany,
[2] Saarbrücken Graduate School of Computer Science,
[2]University of Trieste, Italy

Abstract. We consider the problem of bounding mean first passage times and reachability probabilities for the class of population continuous-time Markov chains, which capture stochastic interactions between groups of identical agents. The quantitative analysis of such models is notoriously difficult since typically neither state-based numerical approaches nor methods based on stochastic sampling give efficient and accurate results. Here, we propose a novel approach that leverages techniques from martingale theory and stochastic processes to generate constraints on the statistical moments of first passage time distributions. These constraints induce a semi-definite program that can be used to compute exact bounds on reachability probabilities and mean first passage times without numerically solving the transient probability distribution of the process or sampling from it. We showcase the method on some test examples and tailor it to models exhibiting multimodality, a class of particularly challenging scenarios from biology.

Keywords: population continuous-time Markov chains · semi-definite programming · exit time distribution · reachability probability · Markov population models

1 Introduction

Population Continuous-Time Markov Chains (PCTMCs) provide a widely used framework to capture stochastic interactions between groups of identical agents. This subclass of Continuous-Time Markov Chains (CTMCs) is used to describe the stochastic dynamics of systems in various domains. Prominent applications are chemical reaction networks in quantitative biology [55], epidemic spreading [46], performance analysis of technical and information systems [11,22] as well as the behavior of collective adaptive systems [9].

For the quantitative analysis of CTMCs, many approaches have been developed, where properties of interest are often expressed in terms of temporal logics such as CSL [2,6,5], MTL [14], and timed-automata specifications [15,41]. In addition, there exist efficient software tools [31,38,17] that can be used to analyze and verify system properties. The computation of reachability probabilities is a central problem in this context.

Popular exact methods for CMTCs rely on numerical approaches that explicitly consider each system state individually. A major problem is that these methods cannot scale in the context of population models with large copy numbers of agents. A popular alternative to tackle this problem is statistical model checking, which is based on stochastic simulation [16]. For PCTMCs arising in the context of chemical reaction networks, trajectories of the process are usually generated using the Stochastic Simulation Algorithm (SSA) [25]. However, since the number of possible interactions grows with the number of agents, stochastic simulations of PCTMCs are time-consuming. Moreover, they are subject to inherent statistical uncertainty and give only statistically estimated bounds.

As an alternative, recent work concentrates on numerical methods that approximate the statistical moments of the system without the need to compute the probability of each state. For groups of identically behaving agents, it is possible to derive systems of differential equations for the evolution of the statistical population moments [10,51,12,21,50,22]. However, as the system of exact moment equations is infinite-dimensional, approximation schemes typically rely on certain assumptions about the underlying probability distribution to truncate it. For example, one might employ a "low dispersion closure" which assumes that higher-order moments are the same as those of a normal distribution [30]. Such approximations are, by nature, ad-hoc and do not come with any guarantees.

Moment-based methods often scale well in terms of population sizes. However, it is not possible to control the effects of the introduced approximations, which in some cases can lead to large errors [50]. This issue reverberates on the application of these methods to compute reachability probabilities and mean first passage times [28,12,13]. Moreover, they can suffer from numerical instabilities, in particular, when the maximum order of the considered moments has to be increased to more appropriately describe the underlying distribution.

Here, we put forward a method based solely on moments that gives *exact bounds* for Mean First Passage Times (MFPTs) and reachability probabilities in PCTMCs. For a set of states, the MFPT within a fixed time-horizon T directly characterizes the probability of reaching that set within T time units. Thus, safe upper and lower bounds on MFPTs can constitute a core component for the verification of properties in PCTMCs. Our approach extends recent work on moment bounds [47,20] and it is based on a martingale formulation of the stopped process that we derive from the exact moment equations. From this formalization, we deduce a set of linear moment constraints from which we derive upper and lower moment bounds using semi-definite programming (SDP). Monotone sequences of both upper and lower bounds can be obtained by increasing the order of the relaxation. Crucially, no closure approximations are introduced. Therefore the bounds are exact up to the numerical accuracy of the SDP solver.

To experimentally validate our method in terms of accuracy and feasibility, we run some tests on examples from biology, leveraging an existing SDP solver and obtaining encouraging results. Comparing with other moment-based methods, our approach is not based on approximations due to closure schemes, thus providing guarantees on the bounds up to the numerical accuracy of the compu-

tations. However, similarly to other moment-based methods, we also found the insurgence of numerical instabilities because moments of higher order tend to span over many orders of magnitude. We ameliorate this problem by considering scaling strategies that reduce such variability. We also extend our approach to deal with PCTMCs exhibiting strong multimodal behavior, due to the presence of populations having low copy numbers. This extension exploits some ideas from hybrid moment closures [34].

In summary, this paper presents the following novel contributions:

- the derivation of moment constraints, based on a martingale formulation, for bounding first passage times and reachability probabilities using a convex programming scheme;
- the extension of this scheme using hybrid moment conditions to systems exhibiting multimodal behavior;
- a scaling strategy for improved robustness during optimization

The paper is structured as follows: Section 2 covers work related to the analysis of first passage times in PCTMCs and recent work on moment bounds. Section 3 introduces the PCTMC framework and its semantics. In Section 4 we derive a martingale from the moment dynamics of a PCTMC. Based on this process, in Section 5 we formulate linear and semi-definite constraints to state a semi-definite program to compute bounds on the MFPT and reachability probabilities. In Section 6, we discuss the practical considerations of the SDP implementation and provide results on a set of case studies. Finally, in Section 7 we provide concluding remarks and directions of future work.

2 Related Work

Considerable effort has been directed at the analysis of first passage time distributions in PCTMCs. Most works can either focus on an explicit state-space analysis [7,43,37,36] or employ approximation techniques for which, in general, no error bounds can be given [49,28,13]. For some model classes such as kinetic proofreading, analytic solutions are possible [43,8,32].

Barzel and Biham [7] propose a recursive scheme that consists of one equation for each state, expressing the average time the system needs to transition from that state to the target state. Kuntz et al. [36] propose to employ moment bounds in a linear programming approach to compute exit time distribution using state-space truncation schemes. In Ref. [37] the authors propose a finite state-space projection scheme to bound first passage time distributions

Hayden et al. [28] use moment closure approximations and Chebychev's inequality to gain an understanding of first passage time dynamics. Schnoerr et al. [49] also employ a moment closure approximation and further approximate threshold functions to derive an approximate first passage time distribution. Bortolussi and Lanciani [13] use a mean-field approximation which is required to reach the target region.

Recently, several groups independently suggested the use of semi-definite optimization for the computation of moment bounds for the limiting distribution [23,19,35,47]. In this approach, the differential equations describing the moment dynamics are set to zero and form linear constraints [3]. Alongside, semi-definite constraints can be placed on the *moment matrices*. These give a semi-definite program that can be solved efficiently.

This approach has been extended to the transient case [20,48]. The approach is similar in both works and is a cornerstone of the MFPT analysis presented here. They differ mainly by the fact that Sakurai and Hori apply a polynomial time-weighting [48], while Dowdy and Barton use an exponential one [20]. We adopt the former approach because it can be naturally adapted to the description of densities over time. The resulting forms can also be adapted to statistical estimation problems [4].

Semi-definite programming has been applied to a wide range of problems, including stochastic processes in the context of financial mathematics [40,33]. For good introductions and overviews of application areas, we refer the reader to Parrilo [45] and, more recently, Lasserre [39].

Particularly relevant for this work is the application of convex optimization to first passage times. Helmes et al. [29] formulated a linear program using the Hausdorff moment conditions to bound moments of the first passage time distribution in Markovian processes. Semi-definite optimization has been successfully applied in financial mathematics by Kashima and Kawai [33], as well as Lasserre et al. [40] to bound prices of exotic options.

3 Preliminaries

A Population Continuous-Time Markov Chain (PCTMC) describes the interactions among a set of agents of n_S types $S_1, \ldots, S_{n_S}$ in a well-stirred reactor.

In the sequel, we will also use other letters than S_i as agent types. Since we assume that all agents are equally distributed in space, we only keep track of the overall copy number of agents for each type. Therefore the state-space is $S \subseteq \mathbb{N}^{n_S}$. The interactions are expressed as *reactions* with a certain gain and loss of agents, given by the non-negative integer vectors $\boldsymbol{\nu}_j^-$ and $\boldsymbol{\nu}_j^+$ for some reaction j, respectively. Such a reaction is denoted as

$$\sum_{i=1}^{n_S} \nu_{ji}^- S_i \xrightarrow{a_j} \sum_{i=1}^{n_S} \nu_{ji}^+ S_i . \tag{1}$$

The reaction rate constant $a_j > 0$ determines the propensity function α_j of the reaction. If just a constant is given, *mass-action* propensities are assumed, where for $\boldsymbol{x} \in S$ we define

$$\alpha_j(\boldsymbol{x}) := a_j \prod_{i=1}^{n_S} \binom{x_i}{\nu_{ji}^-} . \tag{2}$$

This choice of propensity function is natural, since it is proportional to the number of reactant combinations. The system's behavior is described by a stochastic

process $\{\boldsymbol{X}_t\}_{t\geq 0}$. We denote the abundance of a given agent type S_i in $\boldsymbol{X}_t$ by $X_t^{(S_i)}$. The propensity $a_j(\boldsymbol{x})$ gives the infinitesimal probability of a reaction occurring, given a state $\boldsymbol{x}$. That is, for $\boldsymbol{\nu}_j = \boldsymbol{\nu}_j^+ - \boldsymbol{\nu}_j^-$ and a small time step $\Delta t > 0$,

$$\Pr(\boldsymbol{X}_{t+\Delta t} = \boldsymbol{x} + \boldsymbol{\nu}_j \mid \boldsymbol{X}_t = \boldsymbol{x}) = a_j(\boldsymbol{x})\Delta t + o(\Delta t)\,. \tag{3}$$

Therefore, given a system of n_R reactions, the semantics of $\boldsymbol{X}_t$ is given by a continuous-time Markov chain (CTMC) on $\mathbb{S}$ with infinitesimal generator matrix Q with entries

$$Q_{\boldsymbol{x},\boldsymbol{y}} = \begin{cases} -\sum_{j:\boldsymbol{x}+\boldsymbol{\nu}_j=\boldsymbol{y}} a_j(\boldsymbol{x})\,, & \text{if } \boldsymbol{x} \neq \boldsymbol{y}, \\ \sum_{j=1}^{n_R} a_j(\boldsymbol{x})\,, & \text{otherwise.} \end{cases} \tag{4}$$

Accordingly, given an initial distribution on $\mathbb{S}$ the time-evolution of the process' distribution is given by the Kolmogorov forward equation. For a single state, in the context of quantitative biology, it is commonly referred to as the *chemical master equation* (CME)

$$\frac{d\pi}{dt}(\boldsymbol{x}, t) = \sum_{j=1}^{n_R} \left(a_j(\boldsymbol{x} - \boldsymbol{\nu}_j)\pi(\boldsymbol{x} - \boldsymbol{\nu}_j, t) - a_j(\boldsymbol{x})\pi(\boldsymbol{x}, t)\right)\,, \tag{5}$$

where $\pi(\boldsymbol{x}, t) = \Pr(\boldsymbol{X}_t = \boldsymbol{x})$ and $\Pr(\boldsymbol{X}_0 = \boldsymbol{x}) = \pi(\boldsymbol{x}, 0)$.
Consider the following simple PCTMC with non-linear propensities as an example.

Model 1 (Dimerization). *We first examine a simple dimerization model on an unbounded state-space with reactions*

$$\emptyset \xrightarrow{\lambda} M, \quad 2M \xrightarrow{\delta} D$$

and initial condition $X_0^{(M)} = X_0^{(D)} = 0$. *The semantics is given by a CTMC* $\boldsymbol{X}_t = (X_t^{(M)}, X_t^{(D)})^\intercal$, *where* $(S_1, S_2) = (M, D)$. *The reaction propensities according to* (2) *are* $a_1(\boldsymbol{x}) = \lambda$ *and* $a_2(\boldsymbol{x}) = \delta\, x^{(M)}(x^{(M)} - 1)/2$. *The change vectors* $\nu_1^- = (0,0)^\intercal$, $\nu_1^+ = (1,0)^\intercal$, $\nu_2^- = (2,0)^\intercal$, *and* $\nu^+{}_2 = (0,1)^\intercal$. *Consequently,* $\nu_1 = (1, 0)^\intercal$ *and* $\nu_2 = (-2, 1)^\intercal$.

For a state $(x^{(M)}, x^{(D)}) \in \mathbb{N}^2$, where $x^{(M)} \geq 2$, the CME (5) becomes

$$\begin{aligned} \frac{d}{dt}\pi((x^{(M)}, x^{(D)}), t) =& \lambda\pi((x^{(M)} - 1, x^{(D)}), t) \\ &+ \frac{\delta}{2}(x^{(M)} + 2)(x^{(M)} + 1)\pi((x^{(M)} + 2, x^{(D)} - 1), t) \\ &- (\lambda + \frac{\delta}{2}x^{(M)}(x^{(M)} - 1))\pi((x^{(M)}, x^{(D)}), t)\,. \end{aligned}$$

This explicit representation of state probabilities is often not possible, because there are infinitely many states. Usually the state-space is truncated to contain all relevant states [1] or one switches to an approximation such as the mean-field [11].

In this work, we are interested in *first passage times* of such processes. That is the time, the process first enters a set of target states $B \subseteq S$. Naturally, the analysis of first passage times is equivalent to the analysis of times at which the process exits the complement $S \setminus B$. More formally, the first passage time τ for some target set B is defined as the random variable

$$\tau = \inf\{t \geq 0 \mid \boldsymbol{X}_t \in B\}\,. \tag{6}$$

In this example, we are interested in the time at which the number of type M agents exceed some threshold H. With the framework presented in the sequel, one can bound the expected value of this time using semi-definite programming. Further, it is possible to impose a time-horizon T, and find bounds on the probability of $X_t^{(M)} \geq H$ for some $0 \leq t \leq T$. The employed framework is centered around semi-definite relaxations of the generalized moment problem [39]. These require linear constraints on the moments of measures. In the following section, we derive such constraints.

4 Martingale Formulation

Next, we will discuss the ordinary differential equations for the evolution of the statistical moments of the process. The moments over the state-space are then used to derive temporal moments, i.e. moments of measures over both the state-space and the time. This extended description results in a process with the martingale property. This property can be used to formulate linear constraints on the temporal moments and, as a special case, the mean first-passage time. In combination with semi-definite properties of moment matrices, we can formulate mathematical programs that yield upper and lower bounds on mean first passage times.

We start with the description of the *raw moments* dynamics. In particular, a raw moment is

$$\mathsf{E}(\boldsymbol{X}^{\boldsymbol{m}}) = \mathsf{E}\left(\prod_{i=1}^{n_S} X_i^{m_i}\right), \quad \boldsymbol{m} \in \mathbb{N}^{n_S}$$

with respect to some probability measure. The order of a moment $\mathsf{E}(\boldsymbol{X}^{\boldsymbol{m}})$ is given by the sum of its exponents, i.e. $\sum_i m_i$. Note that the notion of expected value can be generalized to any measure μ on a Borel-measurable space $(E, \mathcal{B}(E))$, where the $\boldsymbol{m}$-th raw moment is $\int_E \boldsymbol{x}^{\boldsymbol{m}}\, d\mu(\boldsymbol{x})$. Throughout we assume that moments of arbitrary order remain finite over time, i.e. $\mathsf{E}(|\boldsymbol{X}^{\boldsymbol{m}}|) < \infty$ $\forall t \geq 0$. In Ref. [26] the authors propose a framework to verify this property for a given model.

Let f be a polynomial function, $t \geq 0$. Using the CME (5), we can derive ordinary differential equations (ODEs) describing the dynamics of $\mathsf{E}(f(\boldsymbol{X}_t))$ [21]. Specifically,

$$\frac{d}{dt}\mathsf{E}(f(\boldsymbol{X}_t)) = \sum_{j=1}^{M} \mathsf{E}\left(\left(f(\boldsymbol{X}_t + \boldsymbol{\nu}_j) - f(\boldsymbol{X}_t)\right) a_j(\boldsymbol{X}_t)\right)\,. \tag{7}$$

Let us consider Model 1 as an example and agent type M. Further, let $X_t = X_t^{(M)}$ for ease of exposition. When choosing $f(X_t) = X_t^m$, $m = 1$ and $m = 2$ we obtain two differential equations describing the change of the first two moments of species M, $\mathsf{E}(X_t)$ and $\mathsf{E}[X_t^2]$, respectively.

$$\frac{d}{dt}\mathsf{E}(X_t) = \lambda\mathsf{E}[X_t^0] - 2\delta\,\mathsf{E}[X_t^2] - \mathsf{E}(X_t) \tag{8}$$

$$\frac{d}{dt}\mathsf{E}[X_t^2] = \lambda(2\mathsf{E}(X_t) + 1) - 4\delta\,\mathsf{E}[X_t^3] - 2\mathsf{E}[X_t^2] + \mathsf{E}(X_t). \tag{9}$$

Fixing initial moments, the ODE system describes the moments over time exactly. However, these ODEs cannot be integrated because the system is not closed. The right-hand side for moment $\mathsf{E}(X_t^m)$ always contains $\mathsf{E}[X_t^{m+1}]$. To solve the initial value problem, one typically resorts to ad-hoc approximations of the highest order moments to close the system. Here we do *not* need such approximations because we do not numerically integrate the moment equations. Instead we adopt an approach [20,48] that extends the description of state-space moments to a temporal one.

This is achieved by the introduction of a time-dependent polynomial $\omega(t)$ that is multiplied to (7). An integration by parts on $[0, T]$ yields [20,48]

$$\begin{aligned}&\omega(T)\,\mathsf{E}(f(\boldsymbol{X}_T)) - \omega(0)\,\mathsf{E}(f(\boldsymbol{X}_0)) - \int_0^T \frac{d\omega(t)}{dt}\mathsf{E}(f(\boldsymbol{X}_t))\,dt \\ &= \sum_{j=1}^{R}\int_0^T \omega(t)\,\mathsf{E}((f(\boldsymbol{X}_t + \boldsymbol{v}_j) - f(\boldsymbol{X}_t))\,a_j(\boldsymbol{X}_t))\,dt.\end{aligned} \tag{10}$$

We now want to interchange the order of integration and the summation due to the expected value. To this end, we have to assume the absolute convergence of the integrals. On finite time intervals $[0, T]$ this holds because ω is polynomial and we assumed finite moments for all $t \geq 0$. Interchanging the summation and integral of a monomial $\boldsymbol{x}^{\boldsymbol{m}}$, i.e. pulling all expectation operators outside

$$\int_0^T g(t)\mathsf{E}(\boldsymbol{X}_t^{\boldsymbol{m}})\,dt = \mathsf{E}\left[\int_0^T g(t)\boldsymbol{X}_t^{\boldsymbol{m}}\,dt\right].$$

Hence, we are able to to pull out the expectation operator in (10).

$$\begin{aligned}0 = \omega(T)\mathsf{E}(f(\boldsymbol{X}_T)) - \omega(0)\mathsf{E}(f(\boldsymbol{X}_0)) - \mathsf{E}\left[\int_0^T \frac{d\omega(t)}{dt} f(\boldsymbol{X}_t)\,dt\right] \\ - \sum_{j=1}^{R}\mathsf{E}\left[\int_0^T \omega(t)(f(\boldsymbol{X}_t + \boldsymbol{v}_j) - f(\boldsymbol{X}_t))\,a_j(\boldsymbol{X}_t)\,dt\right],\end{aligned} \tag{11}$$

This gives us the expected value of a time-dependent function of the original process. The function can be viewed as a stochastic process of its own where

the time-horizon T is the index variable. A key property of this process is also illustrated by (11): The process' expected value remains 0, regardless of the choice of T. This martingale property is particularly useful because it can be used to formulate linear constraints on stopping times of the process. Explicitly, we can define this process $\{Z_T\}_{T \geq 0}$ parameterized by the time-weighting ω and polynomial f.

$$\begin{aligned} Z_T := \omega(T) f(\boldsymbol{X}_T) - \omega(0) f(\boldsymbol{X}_0) - \int_0^T \frac{d\omega(t)}{dt} f(\boldsymbol{X}_t)\, dt \\ - \sum_{j=1}^{} \int_0^T \omega(t)(f(\boldsymbol{X}_t + \boldsymbol{\nu}_j) - f(\boldsymbol{X}_t))\, a_j(\boldsymbol{X}_t)\, dt, \end{aligned} \tag{12}$$

A useful choice for f and ω are monomials. When choosing $\omega(t) = t^k$ with $k \in \mathbb{N}$ and $f(\boldsymbol{X}) = \boldsymbol{X}^{\boldsymbol{m}}$ the process takes the form

$$Z_T^{(\boldsymbol{m},k)} = T^k \boldsymbol{X}_T^{\boldsymbol{m}} - 0^k \boldsymbol{X}_0^{\boldsymbol{m}} + \sum_i c_i \int_0^T t^{k_i} \boldsymbol{X}_t^{\boldsymbol{m}_i}\, dt \tag{13}$$

where $(\boldsymbol{m}_i)_i$, $(k_i)_i$, and $(c_i)_i$ are finite sequences resulting from the substitution of f and ω and expansion of (12). This choice allows to naturally characterize the behavior in time and state-space as moments, because the expected value of (13) then becomes a linear form of moments. We will use these as constraints in the semi-definite program used to bound MFPTs.

If we apply this to our previous example (8), letting $m = 1$ and $k = 1$ we obtain the following process for Model 1.

$$Z_T^{(1,1)} = TX_T - \int_0^T X_t\, dt - \lambda \int_0^T t\, dt - 2\delta \int_0^T t X_t\, dt + 2\delta \int_0^T t X_t^2\, dt,$$

where the sequences above are $(m_i)_i = (1, 0, 1, 2)$, $(k_i)_i = (0, 1, 1, 1)$, and $(c_i)_i = (-1, -\lambda, -2\delta, 2\delta)$.

5 Bounds for Mean First Passage Times

We now turn to the analysis of first passage times within some time-bound $T > 0$. Given some subset of the state-space $B \subseteq S$ the first passage time is given by the continuous random variable

$$\tau = \inf\{t \geq 0 \mid \boldsymbol{X}_t \in B\} \wedge T, \tag{14}$$

where $a \wedge b := \min\{a, b\}$. For this work, we only look at threshold hitting times, i.e. we set a threshold H for species S and thus $B = \{\boldsymbol{x} \mid x^{(S)} \geq H\}$. Note, that this framework allows for a more general class of target sets, which are discussed in Section 5.5. In the sequel, we will use τ as a stopping time in our martingale formulation and consider $Z_\tau^{(\boldsymbol{m},k)}$ instead of $Z_T^{(\boldsymbol{m},k)}$. Since (13) defines a martingale, $Z_t^{(\boldsymbol{m},k)}$ remains a martingale by Doob's optional sampling theorem [24]. In particular, this implies that $\mathsf{E}(Z_t^{(\boldsymbol{m},k)}) = 0$ for all moment orders $\boldsymbol{m}$ and degrees k in the weighting function $\omega(t)$.

5.1 Linear Moment Constraints

To simplify our presentation, we fix an initial state $\boldsymbol{x}_0$, i.e. $P(\boldsymbol{X}_0 = \boldsymbol{x}_0) = 1$. Using $\mathsf{E}(Z_\tau^{(\boldsymbol{m},k)}) = 0$ and the form (13) for $Z_\tau^{(\boldsymbol{m},k)}$ yields the following linear constraint on expected values.

$$0 = \mathsf{E}\left[\tau^k \boldsymbol{X}_\tau^{\boldsymbol{m}}\right] - 0^k \boldsymbol{x}_0^{\boldsymbol{m}} + \sum_i c_i \mathsf{E}\left[\int_0^\tau t^{k-1} \boldsymbol{X}_t^{\boldsymbol{m}_i} \, dt\right], \tag{15}$$

where $0^0 = 1$. Hence, we have established a relationship between the process dynamics up to the hitting time via expected values of the time-integrals and the final process state at the hitting time via $\mathsf{E}[\tau^k \boldsymbol{X}_\tau^{\boldsymbol{m}}]$.

For the ease of exposition, we now turn to the analysis of first passage times in one-dimensional processes w.r.t. an upper threshold H. In particular, we will consider moments X^m of a one-dimensional process for $m = 0, 1, 2$ The approach proposed in the sequel, however, can be extended to multi-dimensional processes and more complex target sets B.

Consider again Model 1 and assume that we are interested in the time at which species M exceeds threshold H while fixing the considered time-horizon to $T = 4$. That is, we are interested in the stopping time $\tau = \inf\{t \geq 0 \mid X_t \geq 10\}$. Since the abundance of D does not influence M, we can ignore species D and treat the process as one-dimensional. Figure 1 shows three example trajectories: Two reach an upper threshold $H = 10$, while one reaches the final time-horizon $T = 4$ The figure also illustrates another aspect present in (15). It gives a connection between the terminal distribution, i.e. the distribution of X_τ, and the dynamic behavior up to τ. The statistics at τ are described by a distribution whose moments are represented by the $\mathsf{E}[\tau^k \boldsymbol{X}_\tau^{\boldsymbol{m}}]$ term in (15). This distribution corresponding two moments encompasses both cases of how τ can be reached. In the first case threshold H is reached and the second case the process reaches the time-horizon T. In the following we will define the interplay between these measures more formally.

Therefore we can view (15) as the description of a relationship between two measures [39, Chapter 9.2]:

– *Expected Occupation Measure* ξ supported on $[0, H] \times [0, T]$:

$$\xi(A \times C) := \mathsf{E}\left[\int_{[0,\tau] \cap C} 1_{\in A}(X_t) \, dt\right], \tag{16}$$

– *Exit Location Probability* supported on $(\{H\} \times [0, T]) \cup ([0, H] \times \{T\})$:

$$\nu(A \times C) := \Pr((X_\tau, \tau) \in A \times C), \tag{17}$$

where $A \times C$ is a measurable set, i.e. A and C are elements of the Borel σ-algebras on $[0, H]$ and $[0, T]$, respectively.

Using Figure 1, one can gain an intuition for these two measures. The expected occupation measure is shaded in blue. As the name implies $\xi(A \times C)$

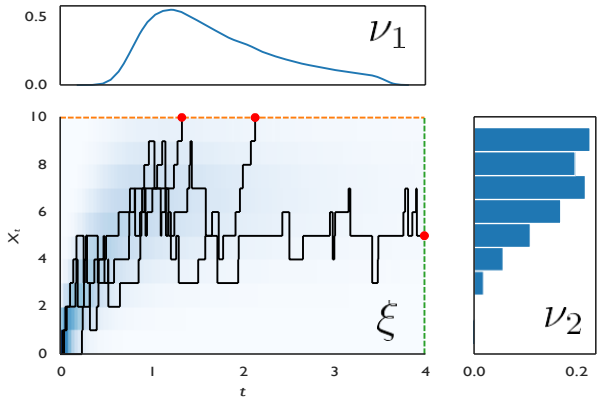


Fig. 1. The relationship between the occupation measure ξ and the exit location probability measures ν_1 and ν_2. The shaded area indicates the structure of the occupation measure. Three example trajectories are additionally plotted with their exit location highlighted. The plots are based on 10,000 sample trajectories.

tells us how much time the process spends in A up to τ restricting to the time instants belonging to C. In particular, $\xi([0, H] \times [0, T]) = \mathsf{E}(\tau)$. The exit location probability ν, while being a two-dimensional distribution, can be viewed as a composition of a density describing the time at which the process reaches H (if it does) and a probability mass function on the states of the process if the time-horizon is reached without exceeding H. We partition the measure ν into ν_1 and ν_2 by conditioning on $\tau = T$. Thus,

$$\nu_1(C) := \Pr(\tau \in C,\ \tau < T) \text{ and } \nu_2(A) := \Pr(X_T \in A,\ \tau = T)$$

and hence $\nu(A \times C) = \nu_1(C) + \nu_2(A)$. To refer to the moments of these measures, we define *partial moments*

$$\mathsf{E}(g(X); f(Y) = y) := \mathsf{E}(g(X) \mid f(Y) = y) \Pr(f(Y) = y),$$

for some polynomial g and some indicator function f. Then

$$\mathsf{E}\left[\tau^k X_\tau^m\right] = T^k \mathsf{E}(X_\tau^m; \tau = T) + H^m \mathsf{E}(\tau^k; \tau < T, X_\tau = H).$$

The partial expectations in terms of ν_1, ν_2

$$\mathsf{E}\left[\tau^k X_\tau^m\right] = \mathsf{E}(X_\tau^m; \tau = T) + \mathsf{E}\left[\tau^k\right]$$

Therefore the linear moment constraints have the form

$$0 = T^k \mathsf{E}(X_\tau^m;\ \tau = T) + H^m \mathsf{E}(\tau^k;\ \tau < T,\ X_\tau = H)$$
$$- 0^k x_0^m + \sum_i c_i \mathsf{E}\left[\int_0^\tau t^k X_t^m \, dt\right]. \tag{18}$$

Next, we consider infinite sequences of partial moments $\boldsymbol{y}_1 = (y_{1k})_{k\in\mathbb{N}}$, $\boldsymbol{y}_2 = (y_{2m})_{m\in\mathbb{N}}$, and $\boldsymbol{z} = (z_{mk})_{(m,k)^T\in\mathbb{N}^2}$ of ν_1, ν_2, and ξ, respectively.

$$y_{1k} := \mathbb{E}\left(\tau^k;\, \tau < T\right), \quad y_{2m} := \mathbb{E}\left(X_\tau^m;\, \tau = T\right), \quad z_{km} := \mathbb{E}\left(\int_0^\tau t^k X_t^m \, dt\right)$$

5.2 Objective

Given the above measures and their corresponding moments, we can now identify the moments we are particularly interested in. We formulate an optimization problem with variables corresponding to the moments defined above. The MFPT is exactly the zeroth moment of ξ,

$$z_{00} = \mathbb{E}\left(\int_0^\tau 1_{\leq H}(X_t)\, dt\right) = \mathbb{E}\left(\tau\right).$$

Therefore z_{00} corresponds to the objective of the optimization problem that gives bounds for the MFPT. Furthermore, we can easily change the objective to the zeroth moment of ν_1,

$$y_{10} = \mathbb{E}\left(\tau^0;\, \tau < T\right) = \Pr(\tau < T).$$

This moment is the probability of reaching threshold H before reaching time-horizon T. Since the target set can be more complex, this formulation can be used to perform model checking on a wide variety of properties.

Moreover, it is possible to formulate objectives not directly corresponding to a raw moment such as the variance [48,19].

5.3 Semi-Deftnite Constraints

The linear constraints alone are not sufficient to identify moment bounds. We further leverage the fact that a necessary condition for a positive measure that the *moment matrices* are positive semi-definite. A matrix $M \in \mathbb{R}^{n\times n}$ is positive semi-definite, denoted by $M \preceq 0$ if and only if

$$\boldsymbol{v}^T M \boldsymbol{v} \geq 0 \quad \forall \boldsymbol{v} \in \mathbb{R}^n.$$

As an example, let us consider a one-dimensional random variable Z with moment sequence $\boldsymbol{z}$. For moment order r, the entries of the $(r+1)\times(r+1)$ moment matrix $M_r(\boldsymbol{x})$ are given by the raw moments. In particular,

$$(M_r(\boldsymbol{z}))_{ij} = z_{i+j-2} = \mathbb{E}\left(Z^{i+j-2}\right)$$

for $i, j \in \mathbb{N}_r$ where $\mathbb{N}_r = 0, 1, \ldots, r$ and the maximum order in the matrix is $2r$. For instance,

$$M_1(\boldsymbol{x}) = \begin{pmatrix} x_0 & x_1 \\ x_1 & x_2 \end{pmatrix} \tag{19}$$

needs to be positive semi-definite. By Sylvester's criterion this means $\det M_1 \geq 0$ and $x_0 \geq 0$. We can easily see that in this case this entails

$$\det M_1 = x_0 x_2 - x_1^2 = \mathbb{E}[X^2] - \mathbb{E}(X)^2 = \mathrm{Var}(X) \geq 0 .$$

This restriction is natural since the variance is always non-negative. This gives us the following restrictions on the moment matrices.

$$M_r(\boldsymbol{z}) \succeq 0, \quad M_r(\boldsymbol{y}_1) \succeq 0, \quad \text{and} \quad M_r(\boldsymbol{y}_2) \succeq 0 \tag{20}$$

for arbitrary orders r, providing a first tranche of moment constraints.

Furthermore, we need to enforce the restriction of the measures ξ, ν_1, and ν_2 to their supports. This can be done, by defining non-negative polynomials on the intended support of the measure. For example, ν_2 has support $[0, H]$. We can now define

$$u_H(t, x) = Hx - x^2, \quad x \in \mathbb{R}$$

as a polynomial that is non-negative on $[0, H]$. Using such polynomials, we can construct *localizing matrices*, which have to be positive semi-definite [39]. Applying u_H to the moment matrix of measure ν_2, i.e. $M_1(\boldsymbol{y}_2)$

$$M(u_H, \boldsymbol{y}_2) = \begin{pmatrix} Hy_{20} - y_{22} & Hy_{21} - y_{23} \\ Hy_{21} - y_{23} & Hy_{22} - y_{24} \end{pmatrix}$$

with the constraint $M_1(u_H, \boldsymbol{y}_2) \succeq 0$, where the application of a polynomial such as u_H to a moment matrix is formally defined for the multidimensional case in Section 5.5. Similarly, let $u_T(t, x) = Tt - t^2$ to restrict ν_1 to $[0, T)$. The expected occupation measure ξ is constrained similarly to its domain $[0, H] \times [0, T]$. This gives us the following restrictions on the moment matrices.

$$M_r(u_T, \boldsymbol{z}) \succeq 0, \quad M_r(u_H, \boldsymbol{z}) \succeq 0, \quad M_r(u_T, \boldsymbol{y}_1) \succeq 0, \quad M_r(u_H, \boldsymbol{y}_2) \succeq 0 . \tag{21}$$

5.4 A Semi-deftnite Program to Bound MFPTs

With the linear constraints given in (15) and the semi-definite constraints (20) and (21) discussed in the previous sections, we can now formulate a semi-definite program (SDP). An SDP is a convex optimization problem over the set of positive semi-definite $n \times n$-matrices $\mathbf{X}$ under linear constraints:

$$\begin{aligned} \min_{X \in \mathbf{X}} \quad & \sum_{i,j} A_{ij}^{(0)} X_{ij} \\ \text{such that } & X \succeq 0 \\ & \sum_{i,j} A_{ij}^{(k)} X \leq b , \quad k = 1, \ldots, m \end{aligned} \tag{22}$$

with constant matrices $A^{(i)} \in \mathbb{R}^{n \times n}$, $i = 0, \ldots, m$ and constants $b_k \in \mathbb{R}$, $k = 1, \ldots, m$ to define a set of m linear constraints. Such a problem is convex and can be solved efficiently [56].

Now we can state the SDP relaxation to the MFPT problem for any order $0 < r < \infty$. With each moment sequence $\boldsymbol{x}$ we associate a sequence proxy variables $\boldsymbol{x}^{\hat{}}$ used in the optimization problem.

$$\begin{aligned} &\min / \max \quad z^{\hat{}}_{00} \\ &\text{such that } M_r(\boldsymbol{z}^{\hat{}}) \succeq 0,\ M_r(u_T, \boldsymbol{z}^{\hat{}}) \succeq 0,\ M_r(u_H, \boldsymbol{z}^{\hat{}}) \succeq 0 \\ &\qquad M_r(\boldsymbol{y}_1^{\hat{}}) \succeq 0, M_r(u_T, \boldsymbol{y}_1^{\hat{}}) \succeq 0 \\ &\qquad M_r(\boldsymbol{y}_2^{\hat{}}) \succeq 0, M_r(u_H, \boldsymbol{y}_2^{\hat{}}) \succeq 0 \\ &\qquad 0 = y^{\hat{}}_{1k} H^m - y^{\hat{}}_{2m} T^k - 0^k x_0^m + \sum_i c_i z^{\hat{}}_{k,m,i} \qquad \forall m, k \end{aligned} \tag{23}$$

This SDP can be compiled to the standard form (22). To this end, the moment matrices can be arranged in a block-diagonal form and the localizing constraints (21) can be encoded by the introduction of new variables and appropriate equality constraints. This transformation can be done automatically using modeling frameworks such as CVXPY [18]. We therefore only give the SDP in the more intuitive format. This problem can be solved using off-the-shelf SDP solvers such as MOSEK [42], CVXOPT [56], or SCS [44].

In principle, we can choose an arbitrarily large order r for the moment matrices and their corresponding constraints, because there are infinitely many moments. In practice, however, the order is bounded by practical issues such as the program size (number of constraints and variables) and numerical issues. These issues are discussed in Section 6 in more detail. Choosing a finite r is a relaxation of the problem since it removes constraints regarding higher-order moments.

5.5 Multi-Dimensional Generalization

For a general multi-dimensional moment sequence $\boldsymbol{y} = (\mathbb{E}(\boldsymbol{X}^{\boldsymbol{m}}))_{\boldsymbol{m} \in \mathbb{N}^{ns}}$, the moment matrix is [39]

$$M_r(\boldsymbol{y})(\boldsymbol{\alpha}, \boldsymbol{\beta}) = y_{\boldsymbol{\alpha}+\boldsymbol{\beta}}, \qquad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n_r$$

where row and column indices, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, are ordered according to the canonical basis

$$\boldsymbol{v}_r(\boldsymbol{x}) = (1, x_1, x_2, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_1 x_n, \ldots, x_1^r, \ldots, x_n^r)^T. \tag{24}$$

Equivalently, $M_r(\boldsymbol{y}) = \mathbb{E}\left[\boldsymbol{v}_r(\boldsymbol{x}) \boldsymbol{v}_r(\boldsymbol{x})^T\right]$. For a moment sequence the semi-definite restriction $M_r(\boldsymbol{y}) \succeq 0$ must hold.

Measures can be restricted to semi-algebraic sets $\{\boldsymbol{x} \in \mathbb{R}^n \mid u_j(\boldsymbol{x}) \geq 0, j = 1, \ldots, m\}$, where u_j, $j = 1, \ldots, m$ are polynomials [39]. This is done by placing restrictions on the localizing matrices. For each polynomial $u_i \in \mathbb{R}[x]$ with coefficient vector $\boldsymbol{u} = \{u_\gamma\}$, i.e. $u(\boldsymbol{x}) = \sum_{\gamma \in \mathbb{N}^n} u_\gamma \boldsymbol{x}^\gamma$, the localizing matrix is

$$M_r(u, \boldsymbol{y})(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{\gamma \in \mathbb{N}^n} u_\gamma y_{\gamma+\boldsymbol{\alpha}+\boldsymbol{\beta}}, \qquad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n_r.$$

Requiring that this matrix is positive semi-definite restricts the measure to $\{\boldsymbol{x} \mid u_i(\boldsymbol{x}) \geq 0\}$. This way we can, for example, restrict the moment sequence $\boldsymbol{y}$ to measures that are positive w.r.t. dimension j. Simply letting $u(\boldsymbol{x}) = x_j$ and requiring $M_1(\boldsymbol{u}, \boldsymbol{y}) \succeq 0$ for $i = 1, \ldots, n_S$ gives us this restriction.

6 Implementation and Evaluation

The implementation of the SDP (23) is straightforward using modeling frameworks and off-the-shelf solvers. However, as noted in previous work [20,47,19,48] on moment-based SDPs the direct implementation of the problem may lead to difficulties for the solver. A source of these is that moments of various orders by nature may differ by many orders of magnitude. A re-scaling of the moments [19,48] such that moments only vary by few orders of magnitude may alleviate this problem. In other scenarios such as the bounding of general transient or steady-state moments, the scaling can be particularly difficult, because the magnitude of moments is generally not known a priori. However, for the MFPT problem, we propose the following moment scaling.

6.1 Moment Scaling

Using the fact that $S \setminus B$ is often finite, it is possible to derive trivial bounds, which can be used to scale moments. If, for example, we have a one-dimensional process X_t with $X_0 = 0$ a.s. and are interested in the hitting time of an upper threshold $H > 0$ until time $T > 0$ for $i, k \in \mathbb{N}$

$$z_{ik} = \mathbb{E}\left[\int_0^\tau t^i X_t^k \, dt\right] \leq \mathbb{E}\left[\int_0^T t^i X_t^k \, dt\right] \leq H^k \int_0^T t^i \, dt = \frac{T^{i+1} H^k}{i+1}.$$

Thus, we fix a scaling vector $\boldsymbol{d}$ with entries $d_{ik} = T^{i+1} H^k$ in the same order as the canonical base vector (24). Using this scaling vector, we can define a scaling matrix $D = \boldsymbol{d}\boldsymbol{d}^\mathsf{T}$. Clearly, $D \succeq 0$. Now we can formulate the optimization (23) over a scaled version $D^{-1} M(\boldsymbol{z}^*)$ instead of $M(\boldsymbol{z}^*)$. The moment matrices of the exit location probabilities are scaled in the same way. Alternatively, one can use approximations such as moment closures or bounds obtained by lower-order relaxations or solve a sequence of problems, incrementally increasing the time-horizon, and adjust the scaling accordingly [20].

In Figure 2 we illustrate the influence the scaling has on the optimization variables. While the unscaled version shows large differences between values, these differences become significantly smaller in the scaled version of the problem.

6.2 Case Studies

We implemented and solved the SDP programs described above using optimization suite MOSEK [42] (version 9.1.2) via the CVXPY interface [18] (version 1.0.24).

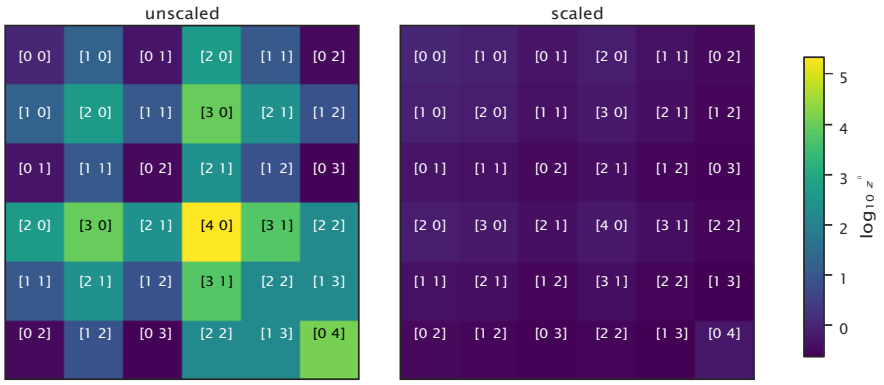
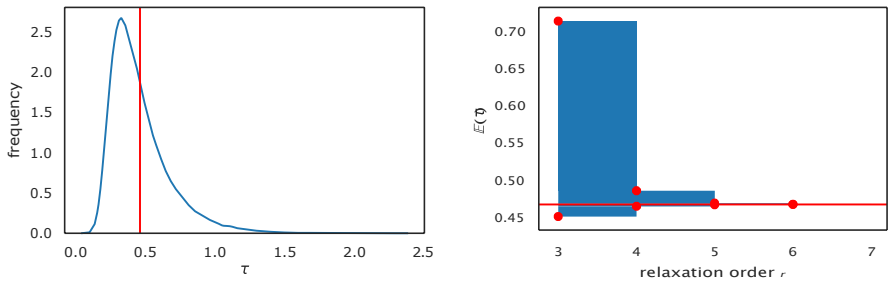


Fig. 2. The unscaled and scaled value the moment matrix proxy variable $M(\boldsymbol{z}^3)$ after optimization using MOSEK. The indices are given along the logarithmic (base 10) values. The unscaled version (left) shows large differences in magnitudes, while on the scaling suppresses these large variations (right). The case study used here is Model 1, with a threshold $H = 25$ for species M and a time-horizon $T = 1$. The relaxation order $r = 2$. Therefore moments of orders up to $2r = 4$ appear.

Dimerization As a first case study, we use Model 1 with parameters $\lambda = 100$ and $\delta = 0.2$. In this model, we are interested in the time at which the number of agents of type M surpasses a threshold of 25 before some time-horizon T, i.e. $\tau = \inf\{t \geq 0 \mid X_t \geq 25\} \wedge T$. First, we set no finite time-horizon T, i.e. $T = \infty$. This is achieved by dropping the moments $\boldsymbol{y}_2$ of measure v_2 in the linear constraints (23). This can be done because the threshold on M makes the state-space finite and therefore the first passage time distribution is a phase-type distribution which possesses finite moments [54, Chapter 7.6]. The empirical FPT distribution based on 100,000 SSA simulations is given in Figure 3a and the bounds, given different moment orders, are given in Figure 3b. As we can see in Figure 3b, the bounds capture the MFPT precisely for orders 5, 6. The difference between upper and lower bound decreases roughly exponentially with increasing relaxation order r. We found that this trend was consistent among the case studies presented here (cf. Figure 5).

Next, we look at first passage times within a finite time-horizon T. In Figure 4a we summarize the bounds obtained for the MFPT over T. While low-order relaxations (light) give rather loose bounds, the bounds are already fairly tight when using $r = 4$. In many cases, hitting probabilities, that is, the probability of reaching the threshold before time T, are of particular interest. This is done by switching the optimization objective in (23) from the mass of the expected occupation measure ξ to the mass of v_1. In terms of moments, the objective changes from z_{00} to y_{10}. The need for such a scenario often arises in the context of model checking, where one might be interested in the probability of a popu-



(a) (b)

Fig. 3. First passage times for Model 1 with $\tau = \inf\{t \geq 0 \mid X_t \geq 10\} \wedge \infty$. The dashed red line denotes the sampled MFPT. (a) The distribution of τ estimated based on 100,000 SSA samples. (b) The bounds based on the SDP in (23) with different moment orders.

lation exceeding a critical threshold. By varying the time-horizon, we are able to recover bounds on the cumulative density $F(t) = \Pr(X_s = H \mid s < t)$ of the first passage time (Fig. 4b).

Finally, we look at turn to the dimer species D that is synthesized by the combination of two monomers M. Here, we look at the time until the agents of type D exceed a threshold of five with a time-horizon $T = 1$. Note that we do not limit the number of M agents. Therefore the analyzed state-space is countably infinite. As in the previous two examples, we observe a roughly exponential decrease in interval size with increasing relaxation order r (cf. Fig. 5 and Table 1).

Parallel Dimerizations As a second study, we consider a 2-dimensional model by combining two independent dimerizations.

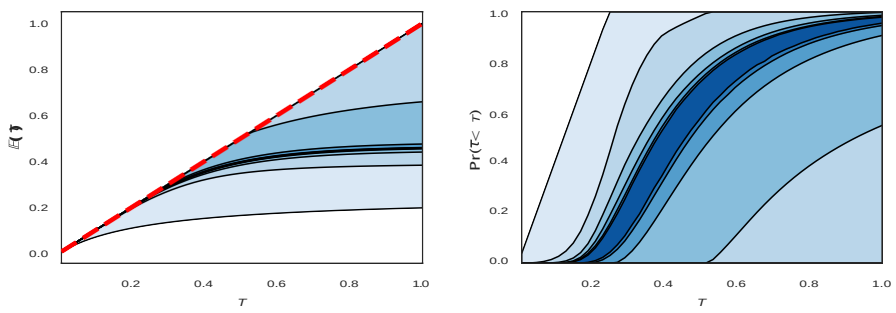
Model 2 (Parallel independent dimerizations).

$$\varnothing \xrightarrow{10^4} M_1, \quad 2M_1 \xrightarrow{0.1} D_1, \quad \varnothing \xrightarrow{10^4} M_2, \quad 2M_2 \xrightarrow{0.1} D_2$$

As a FPT we consider the time at which either M_1 or M_2 surpasses a threshold of 200 or a time-horizon of $T = 10$ is reached, i.e.

$$\tau = \inf\{t \geq 0 \mid X_t^{(M_1)} \geq 200\} \wedge \inf\{t \geq 0 \mid X_t^{(M_2)} \geq 200\} \wedge 10 \,.$$

As before, we ignore the product species D_1 and D_2 since they do not influence τ. The SSA (using $n = 10{,}000$ runs) gives the estimate $\mathsf{E}(\tau) \approx 0.028378$ which is captured tightly by the SDP bounds (cf. Table 1). For higher relaxation orders $r \geq 5$ numerical issues prevented the solution of the corresponding SDPs.



(a) (b)

Fig. 4. First passage times for the dimerization model with $\tau = \inf\{t \geq 0 \mid X_t \geq 25\} \wedge T$. The results for SDP relaxations of orders 1 (light) to 6 (dark) are shown. (a) The bounds on the MFPT for differing time-horizons T. (b) Bounds on the probability to reach the threshold before time T.

6.3 Hybrid Models and Multi-Modal Behavior

The analysis of switching times is a particularly interesting case of FPTs that arises in many contexts. Often mode switching in such systems can be described a modulating Markov process whose switching rates may depend on the system state (e.g. the population sizes). In biological applications, mode switching often describes a change of the DNA state [27,53] and the analysis of switching time distribution is of particular interest [52,7]. In the context of PCTMCs, the state-space of such models can be given as

$$\mathsf{S} = \mathsf{N}^{\tilde{n}_S} \times \{0,1\}^{\hat{n}_S}.$$

This state is modeled by $\hat{n}_S$ population variables with binary domains. Therefore, at each time point, the state of these modulator variables is given by a set of Bernoulli random variables. When considering the moments of such a variable X, clearly $\mathsf{E}(X^m) = \mathsf{E}(X) = \Pr(X = 1)$ for all $m \geq 1$.

We apply a split of variables $\boldsymbol{X}_t$ into the high count part $\tilde{\boldsymbol{X}}_t$ and the binary part $\hat{\boldsymbol{X}}_t$ to the expectations in (7). Similarly, we split $\boldsymbol{\nu}_j$ and with a case distinction over the mode variable, we arrive at a similar result as in [27]:

$$\begin{aligned}\frac{d}{dt}\mathsf{E}\left[\tilde{\boldsymbol{X}}_t^{\boldsymbol{m}} 1_{=\boldsymbol{y}}(\hat{\boldsymbol{X}}_t)\right] = &\sum_{j=1}^{n_R} \mathsf{E}\left[\left(\tilde{\boldsymbol{X}}_t + \tilde{\boldsymbol{\nu}}_j\right)^{\boldsymbol{m}} a_j(\tilde{\boldsymbol{X}}_t, \boldsymbol{y} - \hat{\boldsymbol{\nu}}_j) 1_{=\boldsymbol{y}-\hat{\boldsymbol{\nu}}_j}(\hat{\boldsymbol{X}}_t)\right] \\ &- \sum_{j=1}^{n_R} \mathsf{E}\left[\tilde{\boldsymbol{X}}_t^{\boldsymbol{m}} a_j(\tilde{\boldsymbol{X}}_t, \boldsymbol{y}) 1_{=\boldsymbol{y}}(\hat{\boldsymbol{X}}_t)\right].\end{aligned} \tag{25}$$

Similarly to the general moment case, we can derive a constraint, by multiplying with a time-weighting factor and integrating.

Table 1. MFPT bounds on Models 1, 2, and 3.

Model		Relaxation Order r				
		1	2	3	4	5
Dimerization (Model 1)	lower	0.0909	0.2661	0.2845	0.2867	0.2871
$X^{(D)} \geq 5,\ T = 1$	upper	1.0000	0.3068	0.2932	0.2886	0.2875
Double Dim. (Model 2)	lower	0.0010	0.0250	0.0275	0.0280	0.0280
	upper	10.0000	0.0575	0.0323	0.0299	0.0290
Gene Expression (Model 3)	lower	4.0000	6.0028	6.2207	6.3377	6.3772
	upper	10.7179	6.4619	6.4079	6.4004	6.3835

For simplicity, here we assume $\tilde{n}s = \hat{n}s = 1$. Fixing appropriate sequences $(c_i)_i$, $(m_i)_i$, $(k_i)_i$, and $(y_i)_i$ the constraint has the following form.

$$\sum_{y\in\{0,1\}} H^m \mathbb{E}\left[\tau^k; \hat{X}_\tau = y, \tau < T\right] + T^k \mathbb{E}\left[\tilde{X}_T^m; \hat{X}_T = y, \tau = T\right]$$
$$= 0^k \tilde{x}_0^m 1_{=y}(\hat{x}_0) + \sum_i c_i \mathbb{E}\left[\int_0^\tau t^{k_i} \tilde{X}_t^{m_i}\, dt; \hat{X}_t = y_i\right] \tag{26}$$

This way we can decompose the moment matrices such that for each mode $y \in \{0, 1\}$, we have moment matrices composed of the respective partial moments. To this end, let $z_{k,m}^{(y)}$ be the partial moment w.r.t. $\hat{X} = y$. The moment constraint over the partial moments has a linear structure:

$$0 = y_{1k} H^m - y_{2m} T^k - 0^k x_0^m + \sum_i c_i z_{k_i,m_i}^{(y_i)}. \tag{27}$$

Gene Expression with Negative Feedback As an instance of a multi-modal system, we consider a simple gene expression with self-regulating negative feedback which is a common pattern in many genetic circuits [53].

Model 3 (Negative self-regulated gene expression). *This model consists of a gene state that is either on or off, i.e.* $X^{D_{on}} + X^{D_{off}} = 1,\ \forall t \geq 0$. *Therefore the system has two* modes.

$$D_{on} \xrightarrow{\tau_0} D_{off}, \quad D_{off} \xrightarrow{\tau_1} D_{on}, \quad D_{on} \xrightarrow{\rho} D_{on} + P,$$

$$P \xrightarrow{\delta} \varnothing, \quad P + D_{on} \xrightarrow{\gamma} D_{off}$$

The model parameters are $(\tau_0, \tau_1, \rho, \delta, \gamma) = (10, 10, 2, 0.1, 0.1)$ *and* $_0X^{(D^{off})} = 1$, $X_0^{(P)} = 0$ *a.s.*

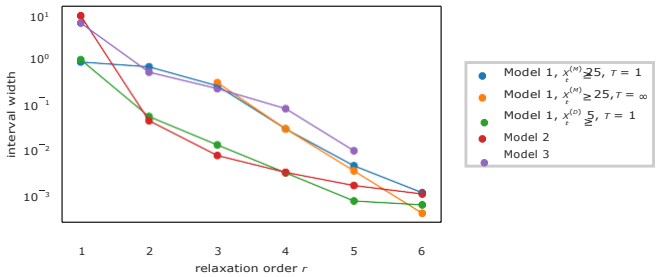


Fig. 5. The interval width, i.e. the difference between upper and lower bound, for different case studies and targeted first passage times against the order r of the SDP relaxation.

As a first passage time we consider

$$\tau = \inf\{t \geq 0 \mid X_t^{(P)} \geq 5\} \wedge 20 .$$

The results are summarized in Table 1. The estimated MFPT based on 100,000 SSA samples is $\mathsf{E}(\tau) \approx 6.37795 \pm 0.02847$ at 99% confidence level. Note that our SDP solution for $r = 5$ yields tighter moment bounds than the statistical estimation.

In Fig. 5 we summarize our results about the decrease of the interval widths for increasing relaxation order r by plotting them on a log-scale. We see an approximately exponential decrease with increasing r. The semi-definite programs above were all solved within at most a few seconds.

7 Conclusion

Numerical methods to compute reachability probabilities and first passage times for continuous-time Markov chains that are based on an exhaustive exploration of the state-space are exact up to numerical precision. Such methods, however, do not scale and cannot be efficiently applied to models with large or infinite state-spaces, an issue exacerbated in population models. Moment-based methods offer an alternative analysis approach for PCTMCs, which scales with the number of different populations in the system but are approximations with little or no control of the error. In this paper, we bridge this gap by proposing a rigorous approach to derive bounds on first passage times and reachability probabilities, leveraging a semi-definite programming formulation based on appropriate moment constraints.

The method we propose is shown to be accurate in several examples. It does, however, suffer, like all moment-based methods, from numerical instabilities in the SDP solver, caused by the fact that moments typically span several orders of magnitude. We proposed a scaling of moments to mitigate this effect. However, the scaling only addresses the moment matrices but not the linear constraints

which still contain values with varying orders of magnitudes. Therefore, we plan as future work to investigate an appropriate scaling for the linear constraints or to redefine the moment constraints (e.g. using an exponential time weighting [20]). Based on this investigation, we expect to make this approach applicable to more problems including, for example, the computation of bounds of rare event probabilities. We also expect that the development of more sophisticated scaling techniques will improve approximate moment-based methods.

Furthermore, moment-based analysis approaches have shown to be successful in a wide range of applications such as optimal control problems or the estimation of densities [39]. We expect that our proposed ideas can be adapted to a wider range of stochastic models such as stochastic hybrid systems, exhibiting partly deterministic dynamics.

Acknowledgements We would like to thank Andreas Karrenbauer for helpful comments on the usage of SDP solvers and Gerrit Großmann for the valuable comments on this manuscript. This work is supported by the DFG project "MULTIMODE", and partially supported by the italian PRIN project "SEDUCE" n. 2017TWRCNB.

References

1. Andreychenko, A., Mikeev, L., Spieler, D., Wolf, V.: Parameter identification for Markov models of biochemical reactions. In: International Conference on Computer Aided Verification. pp. 83 – 98. Springer (2011)
2. Aziz, A., Sanwal, K., Singhal, V., Brayton, R.: Verifying continuous time Markov chains. In: International Conference on Computer Aided Verification. pp. 269 – 276. Springer (1996)
3. Backenköhler, M., Bortolussi, L., Wolf, V.: Moment-based parameter estimation for stochastic reaction networks in equilibrium. IEEE/ACM transactions on computational biology and bioinformatics **15**(4), 1180 – 1192 (2017)
4. Backenköhler, M., Bortolussi, L., Wolf, V.: Control variates for stochastic simulation of chemical reaction networks. In: Bortolussi, L., Sanguinetti, G. (eds.) Computational Methods in Systems Biology. pp. 42 – 59. Springer, Cham (2019)
5. Baier, C., Haverkort, B., Hermanns, H., Katoen, J.P.: Model-checking algorithms for continuous-time Markov chains. IEEE Transactions on software engineering **29**(6), 524 – 541 (2003)
6. Baier, C., Haverkort, B., Hermanns, H., Katoen, J.P.: Model checking continuous-time Markov chains by transient analysis. In: International Conference on Computer Aided Verification. pp. 358 – 372. Springer (2000)
7. Barzel, B., Biham, O.: Calculation of switching times in the genetic toggle switch and other bistable systems. Physical Review E **78**(4), 041919 (2008)
8. Bel, G., Munsky, B., Nemenman, I.: The simplicity of completion time distributions for common complex biochemical processes. Physical biology **7**(1), 016003 (2009)
9. Bernardo, M., De Nicola, R., Hillston, J. (eds.): Formal Methods for the Quantitative Evaluation of Collective Adaptive Systems, Lecture Notes in Computer Science, vol. 9700. Springer International Publishing, Cham (2016)

10. Bogomolov, S., Henzinger, T.A., Podelski, A., Ruess, J., Schilling, C.: Adaptive moment closure for parameter inference of biochemical reaction networks. In: International Conference on Computational Methods in Systems Biology. pp. 77 – 89. Springer (2015)
11. Bortolussi, L., Hillston, J., Latella, D., Massink, M.: Continuous approximation of collective system behaviour: A tutorial. Performance Evaluation **70**(5), 317 – 349 (May 2013)
12. Bortolussi, L., Lanciani, R.: Model checking Markov population models by central limit approximation. In: International Conference on Quantitative Evaluation of Systems. pp. 123 – 138. Springer (2013)
13. Bortolussi, L., Lanciani, R.: Stochastic approximation of global reachability probabilities of Markov population models. In: Computer Performance Engineering - 11th European Workshop, EPEW 2014, Florence, Italy, September 11-12, 2014. Proceedings. pp. 224 – 239 (2014)
14. Chen, T., Diciolla, M., Kwiatkowska, M., Mereacre, A.: Time-bounded verification of CTMCs against real-time specifications. In: International Conference on Formal Modeling and Analysis of Timed Systems. pp. 26 – 42. Springer (2011)
15. Chen, T., Han, T., Katoen, J.P., Mereacre, A.: Quantitative model checking of continuous-time Markov chains against timed automata specifications. In: 2009 24th Annual IEEE Symposium on Logic In Computer Science. pp. 309 – 318. IEEE (2009)
16. David, A., Larsen, K.G., Legay, A., Mikučionis, M., Poulsen, D.B., Sedwards, S.: Statistical model checking for biological systems. International Journal on Software Tools for Technology Transfer **17**(3), 351 – 367 (2015)
17. Dehnert, C., Junges, S., Katoen, J.P., Volk, M.: A storm is coming: A modern probabilistic model checker. In: International Conference on Computer Aided Verification. pp. 592 – 600. Springer (2017)
18. Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. Journal of Machine Learning Research **17**(83), 1 – 5 (2016)
19. Dowdy, G.R., Barton, P.I.: Bounds on stochastic chemical kinetic systems at steady state. The Journal of chemical physics **148**(8), 084106 (2018)
20. Dowdy, G.R., Barton, P.I.: Dynamic bounds on stochastic chemical kinetic systems using semidefinite programming. The Journal of chemical physics **149**(7), 074103 (2018)
21. Engblom, S.: Computing the moments of high dimensional solutions of the master equation. Applied Mathematics and Computation **180**(2), 498 – 515 (2006)
22. Gast, N., Bortolussi, L., Tribastone, M.: Size expansions of mean field approximation: Transient and steady-state analysis. Performance Evaluation **129**, 60 – 80 (2019). https://doi.org/https://doi.org/10.1016/j.peva.2018.09.005
23. Ghusinga, K.R., Vargas-Garcia, C.A., Lamperski, A., Singh, A.: Exact lower and upper bounds on stationary moments in stochastic biochemical systems. Physical biology **14**(4), 04LT01 (2017)
24. Gihman, I., Skorohod, A.: The theory of stochastic processes ii. 1975
25. Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry **81**(25), 2340 – 2361 (1977)
26. Gupta, A., Briat, C., Khammash, M.: A scalable computational framework for establishing long-term behavior of stochastic reaction networks. PLoS Comput Biol **10**(6), e1003669 (2014)
27. Hasenauer, J., Wolf, V., Kazeroonian, A., Theis, F.J.: Method of conditional moments (MCM) for the chemical master equation. Journal of mathematical biology **69**(3), 687 – 735 (2014)

28. Hayden, R.A., Stefanek, A., Bradley, J.T.: Fluid computation of passage-time distributions in large Markov models. Theoretical Computer Science **413**(1), 106 – 141 (2012)
29. Helmes, K., Rohl, S., Stockbridge, R.H.: Computing moments of the exit time distribution for Markov processes by linear programming. Operations Research **49**(4), 516 – 530 (2001)
30. Hespanha, J.: Moment closure for biochemical networks. In: 2008 3rd International Symposium on Communications, Control and Signal Processing. pp. 142 – 147. IEEE (2008)
31. Hinton, A., Kwiatkowska, M., Norman, G., Parker, D.: Prism: A tool for automatic verification of probabilistic systems. In: International Conference on Tools and Algorithms for the Construction and Analysis of Systems. pp. 441 – 444. Springer (2006)
32. Iyer-Biswas, S., Zilman, A.: First-passage processes in cellular biology. Advances in Chemical Physics **160**, 261 – 306 (2016)
33. Kashima, K., Kawai, R.: Polynomial programming approach to weak approximation of lēvy-driven stochastic differential equations with application to option pricing. In: 2009 ICCAS-SICE. pp. 3902 – 3907. IEEE (2009)
34. Kazeroonian, A., Theis, F.J., Hasenauer, J.: Modeling of stochastic biological processes with non-polynomial propensities using non-central conditional moment equation. IFAC Proceedings Volumes **47**(3), 1729 – 1735 (2014)
35. Kuntz, J., Thomas, P., Stan, G.B., Barahona, M.: Rigorous bounds on the stationary distributions of the chemical master equation via mathematical programming. arXiv preprint arXiv:1702.05468 (2017)
36. Kuntz, J., Thomas, P., Stan, G.B., Barahona, M.: Approximation schemes for countably-infinite linear programs with moment bounds. arXiv preprint arXiv:1810.03658 (2018)
37. Kuntz, J., Thomas, P., Stan, G.B., Barahona, M.: The exit time finite state projection scheme: bounding exit distributions and occupation measures of continuous-time Markov chains. SIAM Journal on Scientific Computing **41**(2), A748 – A769 (2019)
38. Kwiatkowska, M., Norman, G., Parker, D.: Prism 4.0: Verification of probabilistic real-time systems. In: International conference on computer aided verification. pp. 585 – 591. Springer (2011)
39. Lasserre, J.B.: Moments, positive polynomials and their applications, vol. 1. World Scientific (2010)
40. Lasserre, J.B., Prieto-Rumeau, T., Zervos, M.: Pricing a class of exotic options via moments and sdp relaxations. Mathematical Finance **16**(3), 469 – 494 (2006)
41. Mikeev, L., Neuhaußer, M.R., Spieler, D., Wolf, V.: On-the-fly verification and optimization of dta-properties for large Markov chains. Formal Methods in System Design **43**(2), 313 – 337 (2013)
42. MOSEK ApS: MOSEK Optimizer API for C 8.1.0.67 (2018), https://docs.mosek.com/8.1/capi/index.html
43. Munsky, B., Nemenman, I., Bel, G.: Specificity and completion time distributions of biochemical processes. The Journal of chemical physics **131**(23), 12B616(2009)
44. O' Donoghue, B., Chu, E., Parikh, N., Boyd, S.: SCS: Splitting conic solver, version 2.1.0. https://github.com/cvxgrp/scs (Nov 2017)
45. Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Mathematical programming **96**(2), 293 – 320 (2003)
46. Porter, M.A., Gleeson, J.P.: Dynamical systems on networks. Frontiers in Applied Dynamical Systems: Reviews and Tutorials **4** (2016)

47. Sakurai, Y., Hori, Y.: A convex approach to steady state moment analysis for stochastic chemical reactions. In: Decision and Control (CDC), 2017 IEEE 56th Annual Conference on. pp. 1206 – 1211. IEEE (2017)
48. Sakurai, Y., Hori, Y.: Bounding transient moments of stochastic chemical reactions. IEEE Control Systems Letters **3**(2), 290 – 295 (2019)
49. Schnoerr, D., Cseke, B., Grima, R., Sanguinetti, G.: Efficient low-order approximation of first-passage time distributions. Phys. Rev. Lett. **119**, 210601 (Nov 2017). https://doi.org/10.1103/PhysRevLett.119.210601
50. Schnoerr, D., Sanguinetti, G., Grima, R.: Comparison of different moment-closure approximations for stochastic chemical kinetics. The Journal of Chemical Physics **143**(18), 185101 (Nov 2015). https://doi.org/10.1063/1.4934990
51. Schnoerr, D., Sanguinetti, G., Grima, R.: Approximation and inference methods for stochastic biochemical kineticsa tutorial review. Journal of Physics A: Mathematical and Theoretical **50**(9), 093001 (Mar 2017). https://doi.org/10.1088/1751-8121/aa54d9
52. Spieler, D., Hahn, E.M., Zhang, L.: Model checking csl for Markov population models. arXiv preprint arXiv:1111.4385 (2011)
53. Stekel, D.J., Jenkins, D.J.: Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression. BMC systems biology **2**(1), 6 (2008)
54. Stewart, W.J.: Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling. Princeton university press (2009)
55. Ullah, M., Wolkenhauer, O.: Stochastic approaches for systems biology. Wiley interdisciplinary reviews. Systems biology and medicine **2**, 385 – 97 (07 2009). https://doi.org/10.1002/wsbm.78
56. Vandenberghe, L.: The cvxopt linear and quadratic cone program solvers. Online: http://cvxopt. org/documentation/coneprog. pdf (2010)