MICHAEL BACKENKÖHLER

ANALYSIS OF MARKOVIAN POPULATION
MODELS

ANALYSIS OF MARKOVIAN POPULATION MODELS

MICHAEL BACKENKÖHLER

DISSERTATION

zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Fakultät für Mathematik und Informatik
der Universität des Saarlandes

Saarbrücken, June 2022

DECLARATION OF ORIGINAL AUTHORSHIP

I hereby declare that this dissertation is my own original work except where otherwise indicated. All data or concepts drawn directly or indirectly from other sources have been correctly acknowledged. This dissertation has not been submitted in its present or similar form to any other academic institution either in Germany or abroad for the award of any other degree.

EIDESSTATTLICHE VERSICHERUNG

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Saarbrücken, June 2022

Michael Backenköhler

ABSTRACT

Short summary of the contents in English. . . a great guide by Kent Beck how to write good abstracts can be found here:

https:
//plg.uwaterloo.ca/~migod/research/beckOOPSLA.html

ZUSAMMENFASSUNG

Kurze Zusammenfassung des Inhaltes in deutscher Sprache. . .

CONTENTS

Part I

PRELIMINARIES

INTRODUCTION

Modelling often is a delicate balance of adequate representation and abstraction. The former is the goal to capture all the relevant behaviours and effects in a model. Abstractions are made for the benefit of both, explainability and facilitating analysis.

Markovian population models (MPMs) 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 [UW09], epidemic spreading [PG16], performance analysis of technical and information systems [Bor+13a; GBT19] as well as the behavior of collective adaptive systems [BDNH16].

In many areas of science, stochastic models of interacting populations can describe systems in which the discrete population sizes evolve stochastically in continuous time. Such problems naturally occur in a wide range of areas such as chemistry [Gil77], systems biology [Wil18; UW11], epidemiology [MS00] as well as queuing systems [Bre03] and finance [Par08].

Interactions between agents, commonly referred to as *reactions*, happen at exponentially distributed random times. Their rate depends on the current system state, i.e. the population sizes. This results in a continuous-time Markov chain semantics [And12].

MPMs are widely used to model the time evolution of complex discrete phenomena in continuous time. Such problems naturally occur in a wide range of areas such as chemistry [Gil77], systems biology [Wil18; UW11], epidemiology [MS00] as well as queuing systems [Bre03] and finance [Par08]. In many applications, an MPM describes the stochastic interaction of populations of agents. The state variables are counts of individual entities of different populations.

Markovian Population Models that are used to describe cellular processes are often subject to inherent stochasticity. The

dynamics of gene expression, for instance, is influenced by single random events (e.g. transcription factor binding) and hence, models that take this randomness into account must monitor discrete molecular counts and reaction events that change these counts. Discrete-state continuous-time Markov chains have successfully been used to describe networks of chemical reactions over time that correspond to the basic events of such processes. The time-evolution of the corresponding probability distribution is given by the chemical master equation, whose numerical solution is extremely challenging because of the enormous size of the underlying state-space.

1.1 ORGANIZATION
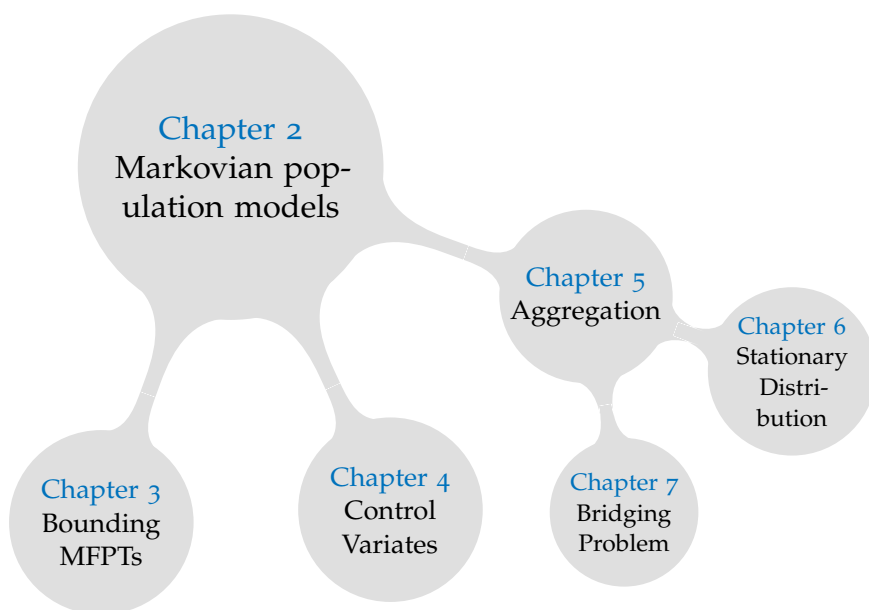
1.2 PREVIOUS PUBLICATIONS

The ideas and much of the presented results have appeared previously in the following publications. As such, most chapters have been published in various conference proceedings. The publications and their respective sections are as indicated below.

- Chapter 3 has with minor differences been published as

Michael Backenköhler, Luca Bortolussi, and Verena Wolf. "Bounding Mean First Passage Times in Population Continuous-Time Markov Chains." In: *17th International Conference on Quantitative Evaluation of SysTems (QEST).* Vol. 12289. LNCS. Springer, 2020, pp. 155–174.

- Chapter 4 has with minor differences been published as

  Michael Backenköhler, Luca Bortolussi, and Verena Wolf. "Control Variates for Stochastic Simulation of Chemical Reaction Networks." In: *17th International Conference on Computational Methods in Systems Biology (CMSB).* Vol. 11773. LNCS. Springer, 2019, pp. 42–59.

- Chapter 6 has with minor differences been published as

  Michael Backenköhler, Luca Bortolussi, Gerrit Großmann, and Verena Wolf. "Abstraction-Guided Truncations for Stationary Distributions of Markovian Population Models." In: *18th International Conference on Quantitative Evaluation of SysTems (QEST).* (2021).

- Chapter 7 has with minor differences been published as

  Michael Backenköhler, Luca Bortolussi, Gerrit Großmann, and Verena Wolf. "Analysis of Markov Jump Processes under Terminal Constraints." In: *27th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS).* Vol. 1265. LNCS. Springer, 2021, pp. 210–229.

Chapter 5 is – in large part – based on this and the preceeding publication.

BACKGROUND

2.1 CONTINUOUS-TIME MARKOV CHAINS

A *stochastic process* is a parameterized collection of random variables $\{X_t\}_{t \in T}$ defined on some probability space $(\Omega, \mathcal{F}, P)$ and takes values in complete metric space $(\mathcal{S}, r)$. In most contexts the index set T models *time*. In a discrete setting $T = \mathbb{N}$, while $T = [0, \infty)$ in the continuous setting, which we consider in this thesis. We observe the values $X(t, \omega)$ for some fixed, but unknown $\omega \in \Omega$. The information of the process up to time t is given by the σ-algebra $\mathcal{F}_t \subset \mathcal{F}$. The increasing family of sigma algebras, i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leqslant t$ is called a *filtration*.

See Øksendal [Øks05] for stoch. processes in general.

More specifically, we study with models having continuous-time Markov chain (CTMC) semantics. A CTMC is a stochastic processes, that takes discrete values $\mathcal{S} = \{s_0, s_1, \dots\}$ over continuous time $T = [0, \infty)$ and fulfills the *Markov property*

$$
\begin{aligned}
&\Pr\left(X_{t_2} = s_2 \mid X_{t_1} = s_1\right) \\
=&\Pr\left(X_{t_2} = s_2 \mid X_{t_1} = s_1, X_{t_0} = s_0\right)
\end{aligned}
\tag{2.1}
$$

for all time points $t_2 \geqslant t_1 \geqslant t_0$ and $s_0, s_1, s_2 \in \mathcal{S}$. Intuitively, this property expresses that the future of the process ($t = t_2$) depends only on the latest condition, i.e. $X_{t_1} = s_1$ and not earlier conditions ($X_{t_0} = x_0$). If further

$$
\Pr\left(X_t = s_1 \mid X_0 = s_0\right) = \Pr\left(X_{t+h} = s_1 \mid X_h = s_0\right) \tag{2.2}
$$

for all $t, h \geqslant 0$ and $s_0, s_1 \in \mathcal{S}$ the chain is *time-homogenous*: The absolute time point is irrelevant, and the dynamics do not change if we shift in time. In this thesis we are only interested in tim-homogenous CTMCs, but most techniques developed should carry over. We define *transition probabilities*

$$
p_{ij}(h) = \Pr(X_h = s_j \mid X_0 = s_i). \tag{2.3}
$$

Accordingly, the *transition matrix* is given by $P(h)_{ij} = p_{ij}(h)$ for all indices i and j. The Chapman-Kolmogorov equation

$$P(s + t) = P(s)P(t) \tag{2.4}$$

follows directly from the law of total probability and the Markov property (2.1). (2.4) directly tells us that the transition probabilities from a semigroup. Studying Markov processes from this direction is a popular approach [EK09].

A standard method of specifying a Markov process, and CTMCs in particular, is is the *generator*. In general, this is an operator A on some class of functions and

$$E\left(f(X_{t+h}) - f(X_t) \mid \mathcal{F}_t\right) = Af(X_t)h + o(h), \tag{2.5}$$

where $\mathcal{F}_t$ is the filtration up to time t. This equation can be interpreted as the requirement, that

$$f(X_t) - f(X_0) - \int_0^t Af(x_s)\,ds$$

is a martingale [Kur81, p. 5]. We will use this in Chapter 3 to derive bounds on mean first-passage times.

In CTMCs the generator is defined by giving the *intensities* or *rates* of transitions. Such a rate $q_{ij} > 0$ between state s_i and s_j implies

$$Pr(X_h = s_j \mid X_0 = s_i) = q_{ij}h + o(h). \tag{2.6}$$

Due to the discrete nature, the generator is a matrix, usually called the Q-matrix,

$$Q = \lim_{h \downarrow 0} \frac{1}{h}\left(P(h) - I\right). \tag{2.7}$$

As such the change of the state probability distribution over time is fully characterized by the *Kolmogorov forward equation*

$$\frac{d}{dt}P(t) = P(t)Q. \tag{2.8}$$

Analogously, the *Kolmogorov backward equation* is

$$\frac{d}{dt}P(t) = QP(t)^\mathsf{T}. \tag{2.9}$$

Distributions in the context of Markov chains are typically in row-vector form

$$\pi(t) := (\pi(x_1, t), \pi(x_2, t), \dots),$$

where we define

$$\pi(x_i, t) := \Pr(X_t = x_i), \quad \forall x_i \in \mathcal{S}, \forall t \geqslant 0.$$

Often, (2.8) and (2.9) are used in the context of an distributions. In this case it makes more sense to right-multiply $\vec{1}$ to these equations such that

$$\frac{d}{dt}\pi(t) = \pi(t)Q \tag{2.10}$$

and

$$\frac{d}{dt}\pi(t) = Q\pi(t)^\top. \tag{2.11}$$

Given some initial distribution

$$\pi_0 := \pi(0),$$

the distribution $\pi(t)$ is given by give simple initial value problems (IVPs) of (2.10).

2.1.1 *Computing Transient Distributions*

Stewart [Ste94] provides a comprehensive overview of solution methods. Here, we only provide a basic overview and intuition relevant for the rest of this thesis.

MATRIX EXPONENTIAL The transition probability matrix $P(t)$ is the solution of (2.8)

$$P(t) = \exp(Qt),$$

where the matrix exponential for a square matrix M

$$\exp(M) := \sum_{k=0}^{\infty} \frac{1}{k!} M^k.$$

Due to the factor $h^k/k!$, the sum of the matrix exponential can be truncated to get an estimate of high quality. In practice, this method is unsuited to many problems, because the factor Q^k becomes incurs a prohibitive cost, especially if the state space is large.

NUMERICAL INTEGRATION The Kolmogorov equations (2.8), (2.9) provide us with an IVP that can be solved numerically. This method scales much better than the matrix exponential method, especially if the whole time-series is of interest. If an initial distribution π_0 is fixed, the ordinary differential equation (ODE) simplify further, such that we only have one equation per state. The drawback to this method is the error inherent to numerical integration schemes.

The matrix exponential is the analytical solution.

UNIFORMIZATION Uniformization is an elegant algorithm to compute transient solutions. Here, the CTMC is transformed into a discrete-time Markov chain (DTMC). Using a uniformization rate

$$\lambda_0 \geqslant \max_i |q_{ii}|$$

the transition probabilities of the DTMC become

$$P_{ij} = \begin{cases} Q_{ij}/\lambda_0\,, & \text{if } i \neq j \\ 1 - \sum_k Q_{ik}/\lambda_0\,, & \text{otherwise} \end{cases}.$$

The transient distribution at t can be obtained, by weighting the k-step probabilities of the DTMC by a Poisson distribution with rate $\lambda_0 t$:

$$\pi(t) = \sum_{k=0}^{\infty} \pi_0 P^k \frac{(\lambda_0 t)^k}{k!} \exp\left(\lambda_0 t\right).$$

Truncation of this series clearly gives an underapproximation.

MONTE CARLO SIMULATION A simple way is estimation using Monte Carlo methods. This entails stochastic simulation of many trajectories of the CTMC. Generating a trajectory is straightforward: Given that the process is in a particular

state s_i the a transition has to be sampled along with the residence time in state s_i. The naive approach is to sample an exponential random variable for each q_{ij} and choose the one firing first. This algorithm can be improved by sampling a reaction directly and sampling the residence time separately. This algorithm will be shown later in the context of MPMs.

2.2 MARKOVIAN POPULATION MODELS

An Markovian population model (MPM) describes the stochastic interactions among agents of n_S distinct types in a well-stirred system. The system is given by a continuous-time stochastic process $\{X_t\}_{t \geqslant 0}$. It models only the number of agents according to their type. Therefore the process takes n_S-dimensional vectors of natural numbers as values, i.e. the state-space is $S \subseteq \mathbb{N}^{n_S}$. By only considering the number of agents, we are neglecting factors such as spatial variations in agent density or other factors influencing interactions. The assumption of all agents being equally distributed in space is called the *well-stirredness* assumptions.

Interactions between agents are expressed as *reactions*. These reactions have associated gains and losses of agents, given by non-negative integer vectors v_j^- and v_j^+ for reaction j, respectively. The overall change by a reaction is given by the vector $v_j = v_j^+ - v_j^-$. A reaction between agents of types $S_1, \ldots, S_{n_S}$ is specified in the following form:

$$\sum_{\ell=1}^{n_S} v_{j\ell}^- S_\ell \xrightarrow{\alpha_j(x)} \sum_{\ell=1}^{n_S} v_{j\ell}^+ S_\ell. \tag{2.12}$$

The propensity function α_j gives the rate of the exponentially distributed firing time of the reaction as a function of the current system state $x \in S$. Thus, for reaction j we have the intensity

$$\Pr\left(X_{t+h} = x + v_j \mid X_t = x\right) = \alpha_j(x) + o(ht). \tag{2.13}$$

In most physical models, *mass-action* propensities are most common. These model combinatorial nature of well-mixed molecules moving randombly through space: In a reaction

$$A + B \xrightarrow{cx^{(A)}x^{(B)}} C$$
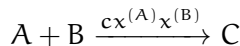
two molecules hit eachother with a probability, proportional to the product of their counts

$$cX_t^{(A)}X_t^{(B)}.$$

In general such rates are given by the product of the number of reactant combinations in x and a *rate constant* c_j, i.e.

$$\alpha_j(x) := c_j \prod_{\ell=1}^{n_S} \binom{x^{(S_\ell)}}{v_{j\ell}^-}. \tag{2.14}$$

This notation avoids conflicts, when we use the subscript for time or as an index.

In this case, we give the rate constant in (2.12) instead of the function α_j. We use the superscript notation $x^{(A)}$ to denote the index corresponding to species A in some vector of length n_S.

According to (2.13) the stochastic process $\{X_t\}_{t \geqslant 0}$ describing the evolution of the population sizes over time t is a continuous-time Markov chain (CTMC). The infinitesimal generator matrix Q has the entries

$$Q_{x,y} = \begin{cases} \sum_{j:x+v_j=y} \alpha_j(x), & \text{if } x \neq y, \\ -\sum_{j=1}^{n_R} \alpha_j(x), & \text{otherwise.} \end{cases} \tag{2.15}$$

Note that in addition mild regularity assumptions are necessary for the existence of a unique CTMC X, such as non-explosiveness [And12]. These assumptions are typically valid for realistic reaction networks. The probability distribution over time is given by an initial value problem. Given an initial state x_0, the distribution

We assume an enumeration of all states in $\mathcal{S}$ with a unique index for each state.

$$\pi(x_i, t) = \Pr(X_t = x_i \mid X_0 = x_0), \quad t \geqslant 0 \tag{2.16}$$

evolves according to the Kolmogorov forward equation

$$\frac{d}{dt}\pi(t) = \pi(t)Q, \tag{2.17}$$

where $\pi(t)$ is an arbitrary vectorization

$$(\pi(x_1, t), \pi(x_2, t), \dots, \pi(x_{|\mathcal{S}|}, t))$$

of the states. (2.16) given for a single state, in the context of quantitative biology, it is commonly referred to as the *chemical master equation* (CME)

$$\frac{d\pi}{dt}(x, t) = \sum_{j=1}^{n_R} \left(\alpha_j(x - v_j)\pi(x - v_j, t) - \alpha_j(x)\pi(x, t) \right). \tag{2.18}$$

EXAMPLE   Consider a birth-death process as a simple example. This model is used to describe a wide variety of phenomena and often constitutes a sub-module of larger models. For example, it represents an M/M/1 queue with service rates being linearly dependent on the queue length. Note that even for this simple model, the state-space is countably infinite.

Model 1 (Birth-Death Process). *The model consists of exponentially distributed arrivals and service times proportional to queue length. It can be expressed using two mass-action reactions:*

$$\varnothing \xrightarrow{\mu} S \quad \text{and} \quad S \xrightarrow{\gamma} \varnothing.$$
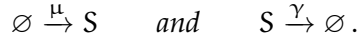
The initial condition $X_0 = 0$ holds with probability one.

For Model 1 the change of probability mass in a single state $x > 0$ is described by expanding (2.18) and

$$\frac{d}{dt}\pi_t(x) = \gamma\pi_t(x-1) + \delta\pi_t(x+1) - (\gamma+\delta)\pi_t(x).$$

$\diamond$

2.3 STATE-SPACE TRUNCATION

A direct solution of (2.18) is usually not possible. If the state-space with non-negligible probability is suitably small, a state space truncation could be performed. That is, (2.18) is integrated on a possibly time-dependent subset $\hat{S}_t \subseteq S$ [HMW09; MK06; Spi14]. Transitions to states, that are not part of this subset are typically re-directed to a introduced sink-state. This state captures the mass "lost" by the approximation and gives the error up to the numerical integration scheme. Munsky and Khammash [MK06] coined the term of finite state projection (FSP) for such a method.

Uniformization can give a lower bound on the error [And+10].

To analyze the stationary distribution (Section 2.6) the redirection scheme needs to be altered [Kun+21b]: Instead of a re-redirection into a sink-state, transitions are redirected in *some fashion* back into the truncation set.

2.4 STOCHASTIC SIMULATIONS

We can generate trajectories of this model using the stochastic simulation algorithm (SSA) (Algorithm 1) [Gil77]. The simulation

algorithm consists of repeatedly evaluating the race condition and jump times induced by (2.15) until some terminal criterion such as a maximum simulation time T is reached (line 2). In particular, the algorithm iteratively chooses a reaction, with a probability that is proportional to its rate given the current state s (line 3). The jump time $t_j - t_{j+1}$ is determined by sampling from an exponential distribution with rate $\sum_i \alpha_i(s)$ (line 4).

Algorithm 1: Sample a trajectory

---

1 $\tau \leftarrow$ empty list, $s \leftarrow$ sample from π_0, $t \leftarrow 0$;

2 **while** $t < T$ **do**

3 $\quad$ $k \leftarrow$ sample reaction i with probability $\alpha_i(s)/\sum_i \alpha_i(s)$;

4 $\quad$ $\delta \sim \mathrm{Exp}\left(\sum_i \alpha_i(s)\right)$;

5 $\quad$ $s \leftarrow s + \nu_k$;

6 $\quad$ $t \leftarrow t + \delta$;

7 $\quad$ $\tau \leftarrow \mathrm{append}(\tau, (s, t))$;

8 **return** τ;

2.5 MOMENT DYNAMICS

Let f be a polynomial function, $t \geqslant 0$. Using the chemical master equation (CME) (2.18), we can derive ODEs describing the dynamics of $E(f(\vec{X}_t))$ [Eng06]. Specifically,

$$\frac{d}{dt} E\left(f(\vec{X}_t)\right) = \sum_{j=1}^{n_R} E\left(\left(f(\vec{X}_t + \vec{v}_j) - f(\vec{X}_t)\right) \alpha_j(\vec{X}_t)\right) . \quad (2.19)$$

Centered or factorial moments (such as the variance) are equivalent via the binomial transform.

This equation is used to analyse (raw) *moments* of the process. A raw moment is

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

with respect to some probability measure. The *order* of a moment $E(\vec{X}^{\vec{m}})$ 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 $\vec{m}$-th raw moment is $\int_E \vec{x}^{\vec{m}} \, d\mu(\vec{x})$. Throughout we assume that moments

of arbitrary order remain finite over time, i.e. $E(|\vec{X}_t^{\vec{m}}|) < \infty$, $t \geqslant 0$. In Gupta, Briat, and Khammash [GBK14] the authors propose a framework to verify this property for a given model.

EXAMPLE   Let us express the dynamics of the first two uncentered moments for Model 1 using (2.19).

$$\frac{\mathrm{d}}{\mathrm{d}t} E\left(X_t\right) = \mu - \gamma E\left(X_t\right)$$
$$\frac{\mathrm{d}}{\mathrm{d}t} E\left(X_t^2\right) = \mu(2E\left(X_t\right) + 1) - \gamma(2E\left(X_t^2\right) - E\left(X_t\right))$$

(2.20)

Setting initial moments these equations give as an IVP, we can solve (see Figure 2.1). This, however, is more an exception than the norm: Unless all ractions have linear or constant rate functions $\alpha_i(\cdot)$, $\forall i$, we would not end up with a closed system of ODEs as in (2.20). To illustrate, let us pretend the reaction ($S \xrightarrow{\gamma} \varnothing$) would become this non-linear reaction

$$2S \xrightarrow{\gamma} \varnothing\,.$$
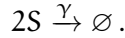
The model is then the same as Model 2.

Accordingly, due to mass-action (2.14)

$$\alpha_2(x) = \gamma(x^2 - x)\,.$$

Therefore the first moment's derivative becomes

$$\frac{\mathrm{d}}{\mathrm{d}t} E\left(X_t\right) = \mu - \gamma\left(E\left(X_t^2\right) - E\left(X_t\right)\right)\,.$$

Note, that now the right-hand side of the derivative in the example depends on the value of the second moment $E\left(X_t^2\right)$.   ◇

If we consider the general expression (2.19) for the moment of order k clearly a term of order $k + 1$ occurs, that does (usually) not cancel out if a propensity function is at least quadratic. Therefore, researches commonly rely on ad-hoc approximations to truncate this infinite system of ODEs [Hes08; SSG15; SSG14]. Unfortunately such schemes have typically no guarantees to converge – or even improve – with increasing truncation order [SSG14] or increasing system size. The only scheme with a convergence guarantee in the system size limit is the *mean-field* approximation [Bor+13b]. Therein zero-covariances are assumed, i.e. the system is truncated at the first oder equations using the approximation $E\left(X_t^2\right) = E\left(X_t\right)^2$.
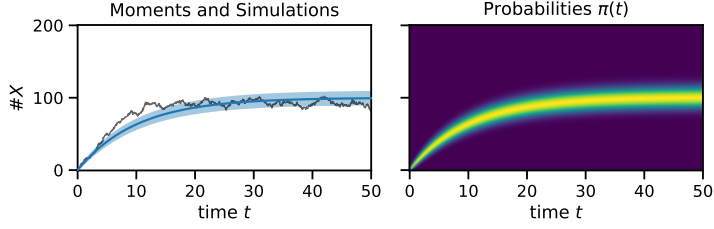
Figure 2.1: The expected value $\pm$ a standard deviation along with a sampled trajectory (left) and the probability distribution over time (right) of Model 1 with $\mu = 10$ and $\gamma = 0.1$.

2.6 STATIONARY DISTRIBUTION

Assuming ergodicity of the underlying chain, a stationary distribution π_∞ is an invariant distribution, namely a fixed point of the Kolmogorov forward equation (2.17). Let π_∞ be the vector description of a stationary distribution. It then satisfies

$$0 = \pi_\infty Q \quad \text{and} \quad 1 = \sum_{x \in S} \pi_\infty(x) \tag{2.21}$$

as a fixed point of the Kolmogorov equation (2.17). Stationary distributions are connected to the *long-run* behavior of an MPM [Day+11], as the system's distribution will converge to the (unique) stationary distribution. The connection of the stationary distribution to the long-run behavior becomes clear when considering the ergodic theorem. For some $A \subseteq S$,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T 1_A(X_t)\, dt = \sum_{x \in A} \pi_\infty(x). \tag{2.22}$$

Thus, the mean occupation time for set A over infinite trajectories is the stationary measure for A. Eq. (2.22) shows that we can assess long-run behavior using the stationary distribution and vice-versa.

EXAMPLE   Returning to the example of Model 1 it is obvious that the state-space is irreducible. Further, we can easily show, that the stationary distribution is Poissonian with rate μ/γ:

$$\pi_\infty(x) = \frac{(\mu/\gamma)^x \exp(-\mu/\gamma)}{x!}.$$

$\diamond$

For simplicity, we assume throughout that the state-space is composed of a single communicating class. Checking ergodicity given a countably infinite number of states is achieved by providing a suitable Foster-Lyapunov function [MT12]. Some automated techniques have been proposed for this task [Day+11; GBK14; MAK14].

2.6.1 *Lyapunov Bounds*

It is well-known that for a CTMC X, ergodicity can be proven by a Lyapunov function $g : \mathcal{S} \to \mathbb{R}_+$ [MT93; Day+11].[1] Given the g, we define its *drift* d as its average infinitesimal change, which is obtained applying the generator Q to g.

$$d(x) = Qg(x) = \sum_{j=1}^{n_R} \alpha_j(x)(g(x + v_j) - g(x)) \qquad (2.23)$$

A Lyapunov function can be used to prove ergodicity of a CTMC: If there is a finite subset $C \subset \mathcal{S}$ such that

$$Qg(x) \leqslant -1, \ \forall x \in \mathcal{S} \setminus C, \qquad (2.24)$$

$$Qg(x) < \infty, \ \forall x \in C, \text{ and} \qquad (2.25)$$

$$\|x\| \to \infty \Rightarrow g(x) \to \infty, \ \text{ where } \ \|x\| = \sum_i x_i, \qquad (2.26)$$

then the chain is non-explosive and ergodic [MAK14; Twe75].

Usually, such a function g grows in all directions on the positive orthant, while its drift $d(x)$ decreases in all directions. More formally, g is characterized by having finite level sets $\{x \in \mathcal{S} \mid g(x) < l\}$ for all $l > 0$. At the same time,

$$\mathcal{C}_{\epsilon_\ell} = \{x \in \mathcal{S} \mid \frac{\epsilon_\ell}{c} d(x) > \epsilon_\ell - 1\} \qquad (2.27)$$

should be finite, where $\infty > c \geqslant \sup_{x \in \mathcal{S}} d(x)$. In this case, $\mathcal{C}_{\epsilon_\ell}$ contains at least $1 - \epsilon_\ell$ of stationary probability mass for any

---

1 The positivity requirement can somewhat be relaxed. The function only needs to be bounded from below since it can be shifted to the positives. In practice this shift can be ignored since it cancels out in the drift (2.23).

$\epsilon_\ell \in (0, 1)$ [Spi14, Thm. 8]. Given that $\mathcal{C}_{\epsilon_\ell}$ is finite, the chain is ergodic and

$$\sum_{x \in \mathcal{C}_{\epsilon_\ell}} \pi(x) > 1 - \epsilon_\ell \tag{2.28}$$

bounding the stationary probability mass contained within $\mathcal{C}_{\epsilon_\ell}$.

2.7   A BRIEF TAXONOMY OF MULTIMODALITY

Multimodality is an overloaded term in the context of reaction network models. Specifically, it can be used to describe the following features.

OPERATIONAL MULTIMODALITY This kind of multimodality characterizes the model behavior directly. Consider, for

example, a gene expression model: The gene state is digital, meaning it is either active or inactive. Depending on this state a protein is either synthesized or not. Therefore the system has distinct *operational modes* which dictate its dynamics. Non-biological examples can be found in the context of broadcasting systems, in which the dynamics change discretely due to messages shared between agents [BHL20].

Naturally, nearly all models change their dynamics, given a change in their state vector. In this aspect this distinction is not wholly strict. It is mainly intended to indicate distinctive changes in the dynamics. In some instances, these can be as obvious as the examples mentioned above. In other cases they may not be as obvious: Consider an epidemics die-out for example, which is a significant change in the operating mode, which is not due to a switch-type reaction.

DISTRIBUTIONAL MULTIMODALITY Here the multimodality refers to the stationary distribution π_∞: It has multiple *modes*, i.e. local maxima. Given an ergodic underlying CTMC this entails, that the system spends most time in distinct regions of the state-space (cf. (2.22)). The *switching* between those distinct region is of interest for both, analysis and control, of such models.

While distributional multimodality implies operational multimodality, the reverse does not hold.

Literature:

- [SG+11]

Part II

MOMENT-BASED METHODS

We use moment properties to bound mean first-passage
times and to improve statistical estimation of different
quantities.

3

BOUNDING MEAN FIRST-PASSAGE TIMES

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 [Azi+96; Bai+00; Bai+03], MTL [Che+11], and timed-automata specifications [Che+09; Mik+13]. In addition, there exist efficient software tools [Hin+06; KNP11; Deh+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 CTMCs 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 [Dav+15]. For MPMs arising in the context of chemical reaction networks, trajectories of the process are usually generated using the SSA [Gil77]. However, since the number of possible interactions grows with the number of agents, stochastic simulations of MPMs 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 [Bog+15; SSG17; BL13; Eng06; SSG15; GBT19]. 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 [Hes08]. 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 [SSG15]. This issue reverberates on the application of these methods to compute reachability probabilities and mean first-passage times [HSBl12; BL13; BL14]. 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 MPMs. For a set of states B and a time-horizon T, the first-passage time (FPT)

$$\tau = \inf\{t \geqslant 0 \mid \vec{X}_t \in B\} \wedge T.$$

This mean of this stopping time $E(\tau)$, i.e. the mean first-passage time, directly characterizes the probability of reaching set B within T time units. Thus, safe upper and lower bounds on MFPTs can constitute a core component for the verification of properties in MPMs. Our approach extends recent work on moment bounds [SH17; DB18b] 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 program (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 computations. 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 MPMs exhibiting strong multimodal behavior, due to the presence of populations having low copy numbers. This extension exploits some ideas from hybrid moment closures [KTH14].

In summary, this chapter presents the following novel contributions:

- the derivation of moment constraints, based on a martingale formulation, for bounding mean 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;

3.1 RELATED WORK

TRUNCATIONS AND ANALYTIC SOLUTIONS   Considerable effort has been directed at the analysis of first-passage time distributions in MPMs. Most works can either focus on an explicit state-space analysis [BB08; MNB09; Kun+19b; Kun+21a] or employ approximation techniques for which, in general, no error bounds can be given [Sch+17; HSBl12; BL14]. For some model classes such as kinetic proofreading, analytic solutions are possible [MNB09; BMN09; IBZ16].

Barzel and Biham [BB08] 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. [Kun+21a] propose to employ moment bounds in a linear programming approach to compute exit time distribution using state-space truncation schemes. In Kuntz et al. [Kun+19b] the authors propose a finite state-space projection scheme to bound first-passage time distributions

MOMENTS APPROXIMATIONS   In Hayden, Stefanek, and Bradley [HSBl12], the authors use moment closure approximations and Chebychev's inequality to gain an understanding of first-passage time dynamics. Schnoerr et al. [Sch+17] also employ a moment closure approximation and further approximate thresh-

old functions to derive an approximate first-passage time distribution. Bortolussi and Lanciani [BL14] use a mean-field approximation which is required to reach the target region.

MOMENT BOUNDS USING OPTIMIZATION   Recently, several groups independently suggested the use of semi-definite optimization for the computation of moment bounds for the limiting distribution [Ghu+17; DB18a; Kun+19a; SH17]. In this approach, the differential equations describing the moment dynamics are set to zero and form linear constraints [BBW18]. 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 [DB18b; SH19]. 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 [SH19] apply a polynomial time-weighting, while Dowdy and Barton [DB18b] use an exponential one. 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 [BBW19].

Semi-definite programming has been applied to a wide range of problems, including stochastic processes in the context of financial mathematics [LPRZ06; KK09]. For good introductions and overviews of application areas, we refer the reader to Parrilo [Par03] and, more recently, Lasserre [Las10].

FIRST PASSAGE TIMES USING OPTIMIZATION   Particularly relevant for this work is the application of convex optimization to FPT. Helmes, Röhl, and Stockbridge [HRS01] formulated a linear program using the Hausdorff moment conditions to bound moments of the FPT distribution in Markovian processes. Semi-definite optimization has been successfully applied in financial mathematics by Kashima and Kawai [KK09], as well as Lasserre, Prieto-Rumeau, and Zervos [LPRZ06] to bound prices of exotic options.

Consider the following simple MPM with non-linear propensities as an example.

Model 2 (Dimerization). *We first examine a simple dimerization model on an unbounded state-space with reactions*

$$\varnothing \xrightarrow{\lambda} M, \quad 2M \xrightarrow{\delta} D$$
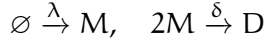
and initial condition $X_0^{(M)} = X_0^{(D)} = 0$. The semantics is given by a CTMC $\vec{X}_t = (X_t^{(M)}, X_t^{(D)})^\mathsf{T}$, where $(S_1, S_2) = (M, D)$. The reaction propensities according to (2.14) are $\alpha_1(\vec{x}) = \lambda$ and $\alpha_2(\vec{x}) = \delta x^{(M)}(x^{(M)} - 1)/2$. The change vectors $v_1^- = (0,0)^\mathsf{T}$, $v_1^+ = (1,0)^\mathsf{T}$, $v_2^- = (2,0)^\mathsf{T}$, and $v_2^+ = (0,1)^\mathsf{T}$. Consequently, $v_1 = (1,0)^\mathsf{T}$ and $v_2 = (-2,1)^\mathsf{T}$.

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 [And+11] or one switches to an approximation such as the mean-field [Bor+13a].

In this work, we are interested in first-passage times (FPTs) of such processes. That is the time, the process first enters a set of target states $B \subseteq S$. Naturally, the analysis of FPTs 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

$$\tau = \inf\{t \geqslant 0 \mid \vec{X}_t \in B\}. \tag{3.1}$$

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 $X_t^{(M)} \geqslant H$ for some $0 \leqslant t \leqslant T$. The employed framework is centered around semi-definite relaxations of the generalized moment problem [Las10]. These require linear constraints on the moments of measures. In the following section, we derive such constraints.

3.3 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.

Let f be a polynomial function, $t \geqslant 0$. Using the CME (2.18), we can derive ODEs describing the dynamics of $E(f(\vec{X}_t))$ [Eng06]. Specifically,

More details on the derivation of the moment ODEs is given in Section 2.5.

$$\frac{d}{dt} E\left(f(\vec{X}_t)\right) = \sum_{j=1}^{n_R} E\left(\left(f(\vec{X}_t + \vec{v_j}) - f(\vec{X}_t)\right) \alpha_j(\vec{X}_t)\right) . \quad (3.2)$$

EXAMPLE   Let us consider Model 2 and agent type M as an example. 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, $E(X_t)$ and $E(X_t^2)$, respectively.

$$\frac{d}{dt} E(X_t) = \lambda E(X_t^0) - 2\delta \left(E(X_t^2) - E(X_t)\right)$$

$$\frac{d}{dt} E(X_t^2) = \lambda(2E(X_t) + 1) - 4\delta \left(E(X_t^3) - 2E(X_t^2) + E(X_t)\right) .$$

$$(3.3)$$

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 $E(X_t^m)$ always contains $E(X_t^{m+1})$.   ◇

To solve the IVP, 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 [DB18b; SH19] that extends the description of state-space moments to a temporal one.

This is achieved by the introduction of a time-dependent polynomial $w(t)$ that is multiplied to (3.2). An integration by parts on $[0, T]$ yields [DB18b; SH19]

$$w(T)\, E\left(f(\vec{X}_T)\right) - w(0)\, E\left(f(\vec{X}_0)\right) - \int_0^T \frac{dw(t)}{dt} E\left(f(\vec{X}_t)\right)\, dt$$

$$= \sum_{j=1}^{n_R} \int_0^T w(t)\, E\left(\left(f(\vec{X}_t + \vec{v}_j) - f(\vec{X}_t)\right)\alpha_j(\vec{X}_t)\right)\, dt .$$

(3.4)

Starting from this equation, it is possible to derive a martingale process, i.e. a process that has an expected value of 0, regardless of time.

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 w is polynomial and we assumed finite moments for all $t \geqslant 0$. Interchanging the summation and integral of a monomial $\vec{x}^{\vec{m}}$, i.e. pulling all expectation operators outside

$$\int_0^T g(t) E\left(\vec{X}_t^{\vec{m}}\right)\, dt = \int_0^T \sum_{x \in S} g(t)\, \Pr(X_s = x) x^m\, dt$$

$$= \int_0^T \int_\Omega g(t) X_s(\omega)^m\, dP(\omega)\, dt$$

$$= \int_\Omega \int_0^T g(t) X_s(\omega)^m\, dt\, dP(\omega)$$

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

Hence, we are able to to pull out the expectation operator in (3.4).

$$0 = w(T) E\left(f(\vec{X}_T)\right) - w(0) E\left(f(\vec{X}_0)\right) - E\left(\int_0^T \frac{dw(t)}{dt} f(\vec{X}_t)\, dt\right)$$

$$- \sum_{j=1}^{n_R} E\left(\int_0^T w(t)(f(\vec{X}_t + \vec{v}_j) - f(\vec{X}_t))\alpha_j(\vec{X}_t)\, dt\right) ,$$

(3.5)

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 (3.5): 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 $\{Z_T\}_{T \geqslant 0}$ parameterized by the time-weighting w and polynomial f.

$$
\begin{aligned}
Z_T := {} & w(T)f(\vec{X}_T) - w(0)f(\vec{X}_0) - \int_0^T \frac{dw(t)}{dt} f(\vec{X}_t)\, dt \\
& - \sum_{j=1}^{n_R} \int_0^T w(t)(f(\vec{X}_t + \vec{v}_j) - f(\vec{X}_t))\alpha_j(\vec{X}_t)\, dt,
\end{aligned}
\tag{3.6}
$$

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

$$
Z_T^{(\vec{m},k)} = T^k \vec{X}_T^{\vec{m}} - 0^k \vec{X}_0^{\vec{m}} + \sum_i c_i \int_0^T t^{k_i} \vec{X}_t^{\vec{m}_i}\, dt
\tag{3.7}
$$

where $(\vec{m}_i)_i$, $(k_i)_i$, and $(c_i)_i$ are finite sequences resulting from the substitution of f and w and expansion of (3.6).

When choosing $w(t) = t^k$ with $k \in \mathbb{N}$ and $f(\vec{X}) = \vec{X}^{\vec{m}}$ this process takes the form

$$
Z_T^{(\vec{m},k)} = T^k \vec{X}_T^{\vec{m}} - 0^k \vec{X}_0^{\vec{m}} + \sum_i c_i \int_0^T t^{k_i} \vec{X}_t^{\vec{m}_i}\, dt
\tag{3.8}
$$

where $(\vec{m}_i)_i$, $(k_i)_i$, and $(c_i)_i$ are finite sequences resulting from the substitution of f and w. This choice allows to naturally characterize the behavior in time and state-space as moments, because the expected value of (3.8) then becomes a linear form of moments. We will use these as constraints in the semi-definite program used to bound MFPTs.

EXAMPLE   If we apply this to our previous example (cf. (3.3)), letting $m = 1$ and $k = 1$ we obtain the following process for Model 2.

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

where the sequences according to (3.8) are $(m_i)_i = (1, 0, 1, 2)$, $(k_i)_i = (0, 1, 1, 1)$, and $(c_i)_i = (-1, -\lambda, -2\delta, 2\delta)$.   ◇

3.4   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

$$\tau = \inf\{t \geqslant 0 \mid \vec{X}_t \in B\} \wedge T$$

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 $B = \{\vec{x} \mid x^{(S)} \geqslant H\}$. Note, that this framework allows for a more general class of target sets, which are discussed in Section 3.5. In the sequel, we will use τ as a stopping time in our martingale formulation and consider $Z_\tau^{(\vec{m},k)}$ instead of $Z_T^{(\vec{m},k)}$. Since (3.8) defines a martingale, $Z_\tau^{(\vec{m},k)}$ remains a martingale by Doob's optional sampling theorem [GS]. In particular, this implies that $E(Z_\tau^{(\vec{m},k)}) = 0$ for all moment orders m and degrees k in the weighting function $w(t)$.

3.4.1   *Linear Moment Constraints*

To simplify our presentation, we fix an initial state $\vec{x}_0$, i.e. $P(\vec{X}_0 = \vec{x}_0) = 1$. Using $E(Z_\tau^{(\vec{m},k)}) = 0$ and the form (3.8) for $Z_\tau^{(\vec{m},k)}$ yields the following linear constraint on expected values.

$$0 = E\left(\tau^k \vec{X}_\tau^{\vec{m}}\right) - 0^k \vec{x}_0^{\vec{m}} + \sum_i c_i E\left(\int_0^\tau t^{k_i} \vec{X}_t^{\vec{m}_i} \, dt\right), \quad (3.9)$$

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 $E\left(\tau^k X_\tau^m\right)$.

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, $m = 0, 1, 2, \ldots$, of a one-dimensional process. The approach proposed

in the sequel, however, can be extended to multi-dimensional processes and more complex target sets B.

Consider again Model 2 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 $\tau = \inf\{t \geqslant 0 \mid X_t \geqslant 10\} \wedge 4$. Since the abundance of D does not influence M, we can ignore species D and treat the process as one-dimensional. Figure 3.1 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 (3.9). 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 $E(\tau^k \vec{X}_\tau^{\vec{m}})$ term in (3.9). 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 (3.9) as the description of a relationship between two measures [Las10, Chapter 9.2]:

EXPECTED OCCUPATION MEASURE ξ describes the expected residence time inside a subset of the state-space and time. As such it is supported on $[0, H] \times [0, T]$:

$$\xi(A \times C) := E \left(\int_{[0,\tau] \cap C} 1_{\in A}(X_t) \, dt \right), \qquad (3.10)$$

EXIT LOCATION PROBABILITY ν gives the state probability associated with the stopping time τ. Therefore it is supported on $(\{H\} \times [0, T]) \cup ([0, H] \times \{T\})$:

$$\nu(A \times C) := \Pr((X_\tau, \tau) \in A \times C), \qquad (3.11)$$

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 3.1, one can gain an intuition for these two measures. The expected occupation measure is shaded in blue. As the name implies $\xi(A \times C)$ tells us how much time the process
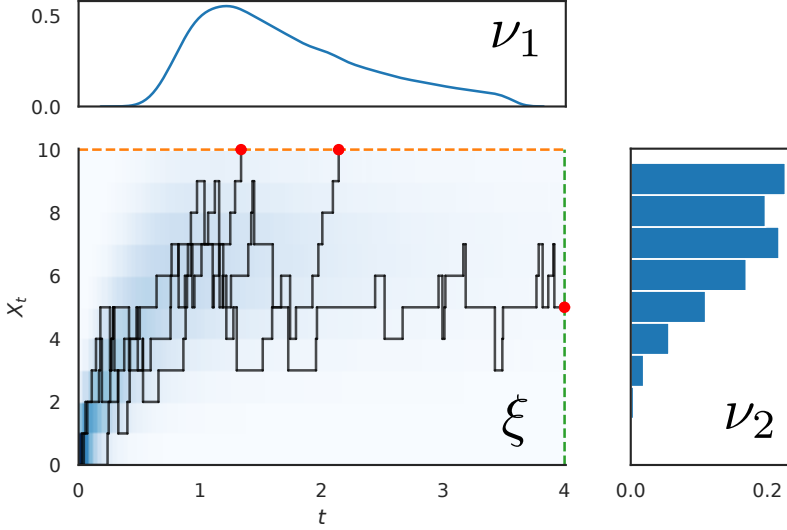
Figure 3.1: The relationship between the occupation measure ξ and the exit location probability measures ν_1 and ν_2. The shaded area indicates the structure of the occupation measure. Three example trajectories are additionally plotted with their exit location highlighted. The plots are based on 10,000 sample trajectories.

spends in A up to τ restricting to the time instants belonging to C. In particular, $\xi([0, H] \times [0, T]) = E(\tau)$. 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,

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

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

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

for some polynomial g and some indicator function f. Then

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

Therefore the linear moment constraints have the form

$$0 = T^k E\left(X_\tau^m; \tau = T\right) + H^m E\left(\tau^k; \tau < T, X_\tau = H\right)$$
$$- 0^k x_0^m + \sum_i c_i E\left(\int_0^\tau t^{k_i} X_t^{m_i}\, dt\right). \tag{3.12}$$

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

$$y_{1k} \coloneqq E\left(\tau^k; \tau < T\right), \quad y_{2m} \coloneqq E\left(X_\tau^m; \tau = T\right),$$

$$z_{km} \coloneqq E\left(\int_0^\tau t^k X_t^m\, dt\right)$$

3.4.2 *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,

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

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,

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

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 [SH19; DB18a].

3.4.3 *Semi-Definite 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 $M \succeq 0$ if and only if

$$\vec{v}^T M \vec{v} \geqslant 0 \quad \forall \vec{v} \in \mathbb{R}^n \,.$$

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

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

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,

$$M_1(\vec{x}) = \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} \tag{3.13}$$

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

$$\det M_1 = x_0 x_2 - x_1^2 = E\left(X^2\right) - E\left(X\right)^2 = \mathrm{Var}(X) \geqslant 0 \,.$$

This restriction is natural since the variance is always non-negative.

$\diamond$

The restriction of the non-negative variance we saw in the example generalizes to moment matrices in form of a positive semi-definite constraint [Par03]. This gives us the following restrictions on the moment matrices.

$$M_r(\vec{z}) \succeq 0, \quad M_r(\vec{y_1}) \succeq 0, \quad \text{and} \quad M_r(\vec{y_2}) \succeq 0 \tag{3.14}$$

This constraint is valid for general positive measures — even if they do not model probabilities.

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.

EXAMPLE   The exit location probability 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 [Las10]. Applying u_H to the moment matrix of measure ν_2, i.e. $M_1(\vec{y}_2)$

$$M_1(u_H, \vec{y}_2) = \begin{bmatrix} Hy_{21} - y_{22} & Hy_{22} - y_{23} \\ Hy_{22} - y_{23} & Hy_{23} - y_{24} \end{bmatrix}$$

with the constraint $M_1(u_H, \vec{y}_2) \succeq 0$, where the application of a polynomial such as u_H to a moment matrix is formally defined for the multidimensional case in Section 3.5. 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.

$$M_r(u_T, \vec{z}) \succeq 0, \quad M_r(u_H, \vec{z}) \succeq 0,$$
$$M_r(u_T, \vec{y_1}) \succeq 0, \quad M_r(u_H, \vec{y_2}) \succeq 0. \tag{3.15}$$

$\diamond$

3.5   MULTI-DIMENSIONAL GENERALIZATION

For a general multi-dimensional moment sequence

$$\vec{y} = \left(E\left(\vec{X}^{\vec{m}} \right) \right)_{\vec{m} \in \mathbb{N}^{n_s}},$$

the moment matrix is [Las10]

$$M_r(\vec{y})(\vec{\alpha}, \vec{\beta}) = y_{\vec{\alpha} + \vec{\beta}}, \quad \forall \vec{\alpha}, \vec{\beta} \in \mathbb{N}_r^n$$

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

$$\vec{v}_r(\vec{x}) = (1, x_1, x_2, \dots, x_n, x_1^2, x_1 x_2, \dots, x_1 x_n, \dots, x_1^r, \dots, x_n^r)^\top. \tag{3.16}$$

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

Measures can be restricted to semi-algebraic sets

$$\{\vec{x} \in \mathbb{R}^n \mid u_j(\vec{x}) \geqslant 0, j = 1, \ldots, m\},$$

where u_j, $j = 1, \ldots, m$ are polynomials [Las10]. This is done by placing restrictions on the localizing matrices. For each polynomial $u_i \in \mathbb{R}[x]$ with coefficient vector $\vec{u} = \{u_{\vec{\gamma}}\}$, i.e. $u(\vec{x}) = \sum_{\vec{\gamma} \in \mathbb{N}^n} u_{\vec{\gamma}} \vec{x}^{\vec{\gamma}}$, the localizing matrix is

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

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

3.5.1 *A Semi-Definite Program to Bound MFPTs*

With the linear constraints given in (3.9) and the semi-definite constraints (3.14) and (3.15) discussed in the previous sections, we can now formulate a semi-definite program (SDP) for any relaxation order $0 < r < \infty$. An SDP is a convex optimization problem over the set of positive semi-definite $n \times n$-matrices X under linear constraints:

$$\begin{aligned} \min_{X \in \mathcal{X}} \quad & \sum_{i,j} A_{ij}^{(0)} X_{ij} \\ \text{such that} \quad & X \succeq 0 \\ & \sum_{i,j} A_{ij}^{(k)} X_{ij} \leqslant b_k, \quad k = 1, \ldots, m \end{aligned} \qquad (3.17)$$

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 using off-the-shelf solvers [Van10].

With each moment sequence $\vec{x}$ we associate a sequence proxy variables $\vec{x'}$ used in the optimization problem.

$$
\begin{aligned}
\min/\max \quad & z'_{00} \\
\text{such that} \quad & M_r(\vec{z'}) \succeq 0, M_r(u_T, \vec{z'}) \succeq 0, M_r(u_H, \vec{z'}) \succeq 0 \\
& M_r(\vec{y'_1}) \succeq 0, M_r(u_T, \vec{y'_1}) \succeq 0 \\
& M_r(\vec{y'_2}) \succeq 0, M_r(u_H, \vec{y'_2}) \succeq 0 \\
& 0 = y'_{1k} H^m - y'_{2m} T^k - 0^k x_0^m + \sum_i c_i z'_{k_i m_i}, \quad \forall m, k
\end{aligned}
$$

$$(3.18)$$

This SDP can be compiled to the canonical form. To this end, the moment matrices can be arranged in a block-diagonal form and the localizing constraints (3.15) 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 [DB16]. 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 [MOSEK ApS18], CVXOPT [Van10], or SCS [O'D+17].

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 3.6 in more detail. Choosing a finite r is a relaxation of the problem since it removes constraints regarding higher-order moments.

3.6   IMPLEMENTATION AND EVALUATION

The implementation of the SDP (3.18) is straightforward using modeling frameworks and off-the-shelf solvers. However, as noted in previous work [DB18b; SH17; DB18a; SH19] 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 [DB18a; SH19] 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. In the context of MFPTs with a finite time-horizon moments are trivially bounded.

3.6.1 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 $i, k \in \mathbb{N}$

$$z_{ik} = E\left(\int_0^\tau t^i X_t^k \, dt\right) \leqslant E\left(\int_0^T t^i X_t^k \, dt\right) \leqslant H^k \int_0^T t^i \, dt = \frac{T^{i+1} H^k}{i+1}.$$

Thus, we fix a scaling vector $\vec{d}$ with entries $d_{ik} = T^{i+1} H^k$ in the same order as the canonical base vector (3.16). Using this scaling vector, we can define a scaling matrix $D = \vec{d}\vec{d}^T$. Clearly, $D \succeq 0$. Now we can formulate the optimization (3.18) over a scaled version $D^{-1} M(\vec{z'})$ instead of $M(\vec{z'})$. The moment matrices of the exit location probabilities are scaled in the same way. Alternatively, one could 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 (suggested in [DB18b]).

These scaling strategies have not been evluated so far.

In Figure 3.2 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.

3.6.2 Case Studies

We implemented and solved the SDP programs described above using optimization suite MOSEK [MOSEK ApS18] (version 9.1.2) via the CVXPY interface [DB16] (version 1.0.24).
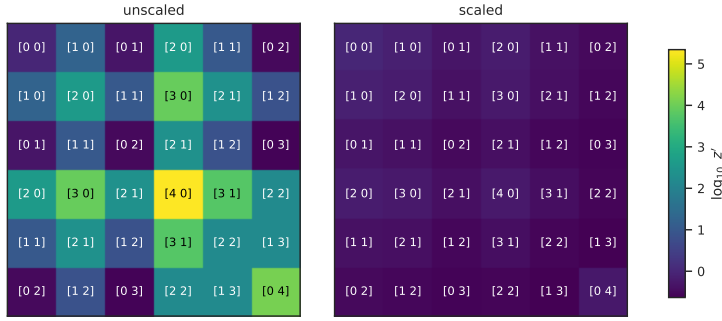
Figure 3.2: The unscaled and scaled value of the moment matrix proxy variable $M(\vec{z'})$ after optimization using MOSEK. The indices are given along the logarithmic (base 10) values. The unscaled version (left) shows large differences in magnitudes, while on the scaling suppresses these large variations (right). The case study used here is Model 2, 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.

3.6.2.1 *Dimerization*

As a first case study, we use Model 2 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.

$$\tau = \inf\{t \geqslant 0 \mid X_t \geqslant 25\} \wedge T.$$

First, we set no finite time-horizon T, i.e. $T = \infty$. This is achieved by dropping the moments $\vec{y}_2$ of measure ν_2 in the linear constraints (3.18). 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 [Ste09, Chapter 7.6]. The empirical FPT distribution based on 100,000 SSA simulations is given in Figure 3.3a and the bounds, given different moment orders, are given in Figure 3.3b. As we can see in Figure 3.3b, the bounds capture the MFPT precisely for orders 5, 6. The difference between upper and lower
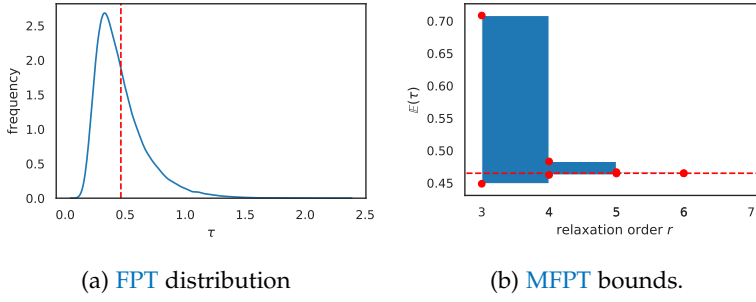
(a) FPT distribution

(b) MFPT bounds.

Figure 3.3: First-passage time characterisitics for Model 2 with $\tau = \inf\{t \geqslant 0 \mid X_t \geqslant 10\} \wedge \infty$. The dashed red line denotes the sampled MFPT based on 100,000 SSA samples. Bounds are based on the SDP (3.18) with different moment orders

bound decreases roughly exponentially with increasing relaxation order r. We found that this trend was consistent among the case studies presented here (cf. Figure 3.5).

Next, we look at first passage times within a finite time-horizon T. In Figure 3.4a 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 (3.18) from the mass of the expected occupation measure ξ to the mass of ν_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 population 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 (Figure 3.4b).

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
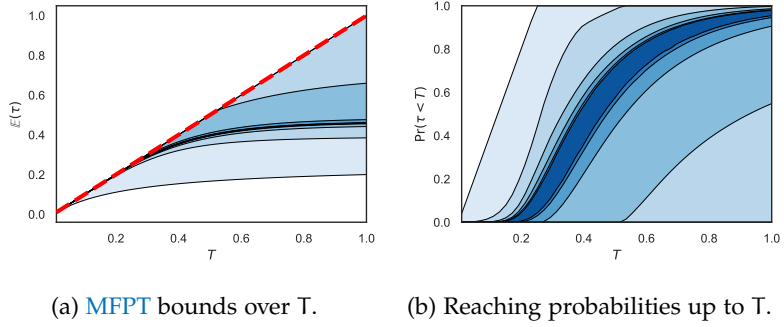
(a) MFPT bounds over T.

(b) Reaching probabilities up to T.

Figure 3.4: MFPTs and reaching probability bounds for the dimeriza-
tion model with $\tau = \inf\{t \geqslant 0 \mid X_t \geqslant 25\} \wedge T$ and varying T.
The results for SDP relaxations of orders 1 (light) to 6 (dark)
are shown.

infinite. As in the previous two examples, we observe a roughly
exponential decrease in interval size with increasing relaxation
order r (cf. Figure 3.5 and Table 3.1).

3.6.2.2 *Parallel Dimerizations*

As a second study, we consider a 2-dimensional model by com-
bining two independent dimerizations.

Model 3 (Parallel independent dimerizations)**.**

$$\varnothing \xrightarrow{10^4} M_1, \quad 2M_1 \xrightarrow{0.1} D_1, \quad \varnothing \xrightarrow{10^4} M_2, \quad 2M_2 \xrightarrow{0.1} D_2$$
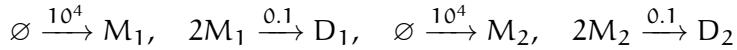
As a FPT we consider the time at which either M_1 or M_2
surpasses a threshold of 200 or a time-horizon of $T = 10$ is
reached, i.e.

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

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 $E(\tau) \approx 0.028378$ which is captured tightly by the SDP
bounds (cf. Table 3.1). For higher relaxation orders $r \geqslant 5$ numeri-
cal issues prevented the solution of the corresponding SDPs.

3.6.3 *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 [Has+14; SJo8] and the analysis of switching time distribution is of particular interest [SHZ11; BB08]. In the context of MPMs, typically the state-space $\mathcal{S} = \mathbb{N}^{\tilde{n}_S} \times \{0, 1\}^{\hat{n}_S}$. 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 $E(X^m) = E(X) = \Pr(X = 1)$ for all $m \geqslant 1$.

We apply a split of variables $\vec{X}_t$ into the high count part $\vec{\tilde{X}}_t$ and the binary part $\vec{\hat{X}}_t$ to the expectations in (3.2). Similarly, we split $\vec{v}_j$ and with a case distinction over the mode variable, we arrive at a similar result as in [Has+14]:

$$
\frac{d}{dt} E\left(\vec{\tilde{X}}_t^{\vec{m}} 1_{=\vec{y}}(\vec{\hat{X}}_t)\right)
$$

$$
= \sum_{j=1}^{n_R} E\left(\left(\left(\vec{\tilde{X}}_t + \vec{\tilde{v}}_j\right)^{\vec{m}} 1_{=\vec{y}}(\vec{\hat{X}}_t + \vec{\hat{v}}_j) - \vec{\tilde{X}}_t^{\vec{m}} 1_{=\vec{y}}(\vec{\hat{X}}_t)\right) \alpha(\vec{X}_t)\right)
$$

$$
= \sum_{j=1}^{n_R} E\left(\left(\vec{\tilde{X}}_t + \vec{\tilde{v}}_j\right)^{\vec{m}} \alpha_j(\vec{X}_t, \vec{y} - \vec{\hat{v}}_j) 1_{=\vec{y} - \vec{\hat{v}}_j}(\vec{\hat{X}}_t)\right)
$$

$$
- \sum_{j=1}^{n_R} E\left(\vec{\tilde{X}}_t^{\vec{m}} \alpha_j(\vec{X}_t, \vec{y}) 1_{=\vec{y}}(\vec{\hat{X}}_t)\right) .
$$

$$(3.19)$$

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

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.   *We fix* $0^0 = 1$.

Table 3.1: MFPT bounds on Model 2, Model 3, Model 4.

		Relaxation Order r				
		1	2	3	4	5
Model 2	lower	0.0909	0.2661	0.2845	0.2867	0.2871
	upper	1.0000	0.3068	0.2932	0.2886	0.2875
Model 3	lower	0.0010	0.0250	0.0275	0.0280	0.0280
	upper	10.0000	0.0575	0.0323	0.0299	0.0290
Model 4	lower	4.0000	6.0028	6.2207	6.3377	6.3772
	upper	10.7179	6.4619	6.4079	6.4004	6.3835

$$
\sum_{y \in \{0,1\}} H^m E\left(\tau^k; \hat{X}_\tau = y, \tau < T\right)
$$
$$
+ T^k E\left(\tilde{X}_T^m; \hat{X}_T = y, \tau = T\right) \tag{3.20}
$$
$$
= 0^k \tilde{x}_0^m 1_{=y}(\hat{x}_0) + \sum_i c_i E\left(\int_0^\tau t^{k_i} \tilde{X}_t^{m_i} \, dt; \hat{X}_t = y_i\right)
$$

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 $z_m^{(y)}$ be the partial moment w.r.t. $\hat{X} = y$. The moment constraint over the partial moments has a linear structure:

$$
0 = y_{1k} H^m - y_{2m} T^k - 0^k x_0^m + \sum_i c_i z_{k_i m_i}^{(y_i)}. \tag{3.21}
$$

3.6.3.1 *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 [SJ08].

Model 4 (Negative self-regulated gene expression). *This model consists of a gene state that is either on or off, i.e.* $X_t^{D_{on}} + X_t^{D_{off}} = 1$, *$\forall t \geqslant 0$. Therefore the system has two* modes.

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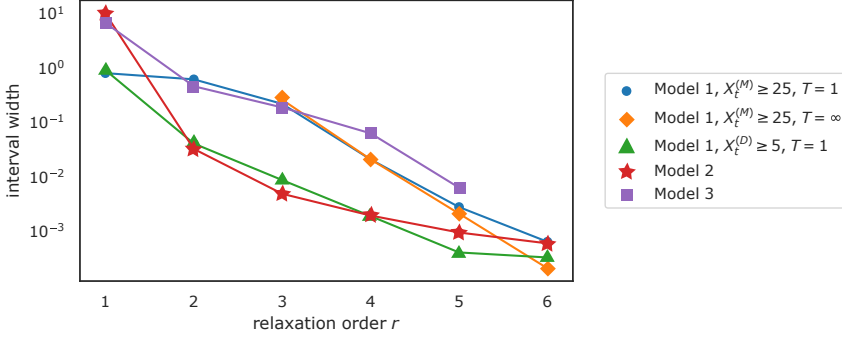$$

Figure 3.5: 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.

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

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

As a first passage time we consider

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

The results are summarized in Table 3.1. The estimated MFPT
based on 100,000 SSA samples is $E(\tau) \approx 6.37795 \pm 0.02847$ at
99% confidence level. Note that our SDP solution for $r = 5$ yields
tighter moment bounds than the statistical estimation.

In Figure 3.5 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 de-
crease with increasing r. The SDPs above were all solved within
at most a few seconds.

3.7 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 ex-
act 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 MPMs, which scales with the number of different populations in the system but are approximations with little or no control of the error. In this work, 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 [DB18b]). 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 [Las10]. 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.

4

CONTROL VARIATES FOR MONTE-CARLO ESTIMATION

Analysis approaches based on sampling, such as the stochastic simulation algorithm (SSA) [Gil77], can be applied independent of the size of the model's state-space. However, statistical approaches are costly since a large number of simulation runs is necessary to reduce the statistical inaccuracy of estimators. This problem is particularly severe if reactions occur on multiple time scales or if the event of interest is rare. A particularly popular technique to speed up simulations is τ-leaping which applies multiple reactions in one step of the simulation. However, such multi-step simulations rely on certain assumptions about the number of reactions in a certain time interval. These assumptions are typically only approximately fulfilled and therefore introduce approximation errors on top of the statistical uncertainty of the considered point estimators.

Moment-based techniques offer a fast approximation of the statistical moments of the model. The exact moment dynamics can be expressed as an infinite-dimensional system of ODEs, which cannot be directly integrated for a transient analysis. Hence, ad-hoc approximations need to be introduced, expressing higher order moments as functions of lower-order ones [AKS13; Eng06]. However, moment-based approaches rely on assumptions about the dynamics that are often not even approximately fulfilled and may lead to high approximation errors. Recently, equations expressing the moment dynamics have also been used as constraints for parameter estimation [BBW18] and for computing moment bounds using semi-definite programming [DB18b; Ghu+17].

In this work, we propose a combination of such moment constraints with the SSA approach. Specifically, we interpret these constraints as random variables that are correlated with the estimators of interest usually given as functions of chemical population variables. These constraints can be used as (linear) control variates (CVs) in order to improve the final estimate and reduce

its variance [LMW82; Sze03]. The method is easy on an intuitive level: If a control variate is positively correlated with the function to be estimated then we can use the estimate of the variate to adjust the target estimate.

The incorporation of control variates into the SSA introduces additional simulation costs for the calculation of the constraint values. These values are integrals over time, which we accumulate based on the piece-wise constant trajectories. This introduces a trade-off between the variance reduction that is achieved by using control variates versus the increased simulation cost. This trade-off is expressed as the product of the variance reduction ratio and the cost increase ratio.

For a good trade-off, it is crucial to find an appropriate set of control variates. Here we propose a class of constraints which is parameterized by a moment vector and a weighting parameter, resulting in infinitely many choices. We present an algorithm that samples from the set of all constraints and proceeds to remove constraints that are either only weakly correlated with the target function or are redundant in combination with other constraints.

In a case study, we explore different variants of this algorithm both in terms of generating the initial constraint set and of removing weak or redundant constraints. We find that the algorithm's efficiency is superior to a standard estimation procedure using stochastic simulation alone in almost all cases.

Although in this work we focus on estimating first order moments at fixed time points, the proposed approach can in principle deal with any property that can be expressed in terms of expected values such as probabilities of complex path properties. Another advantage of our technique is that an increased efficiency is achieved without the price of an additional approximation error as it is the case for methods based on moment approximations or multi-step simulations.

4.1   RELATED WORK

STATE-SPACE TRUNCATIONS   If the state-space is finite and small enough one can deal with the underlying Markov chain directly. But there are also cases where the transient distribution has an infinitely large support and one can still deal with explicit

state probabilities. To this end, one can fix a finite state-space, that should contain most of the probability [MK06]. Refinements of the method work dynamically and adjust the state-space according to the transient distributions [And+11; HMW09; Mat+10].

MOMENT APPROXIMATIONS   On the other end of the spectrum there are mean-field approximations, which model the mean densities faithfully in the system size limit [Bor+13b]. In between there are techniques such as moment closure [SH06], that not only consider the mean, but also the variance and other higher order moments. These methods depend on ad-hoc approximations of higher order moments to close the ODE system given by the moment equations. Yet another class of methods approximate molecular counts continuously and approximate the dynamics in such a continuous space, e.g. the system size expansion [VK92] and the chemical Langevin equation [Gil00].

While the moment closure method uses ad-hoc approximations for high order moments to facilitate numerical integration, they can be avoided in some contexts. For the equilibrium distribution, for example, the time-derivative of all moments is equal to zero. This directly yields constraints that have been used for parameter estimation at steady-state [BBW18] and bounding moments of the equilibrium distribution using semi-definite programming [GLS18; Ghu+17; Kun+19a]. The latter technique of bounding moments has been successfully adapted in the context of transient analysis [DB18b; SH17; SH19]. We adapt the constraints proposed in these works to improve statistical estimations via stochastic simulation (cf. Section 4.2).

MONTE-CARLO SIMULATION   While the above techniques give a deterministic output, stochastic simulation generates single executions of the stochastic process [Gil77]. This necessitates accumulating large numbers of simulation runs to estimate quantities. This adds a significant computational burden. Consequently, considerable effort has been directed at lowering this cost. A prominent technique is τ-leaping [Gil01], which in one step performs multiple instead of only a single reaction. Another approach is to find approximations that are specific to the problem at hand,

such as approximations based on time-scale separations [CGP05; BMS15].

Recently, multilevel Monte Carlo methods have been applied in to time-inhomogenous MPM [AY18]. In this techniques estimates are combined using estimates of different approximation levels.

The most prominent application of a variance reduction technique in the context of stochastic reaction networks is importance sampling [KM08]. This technique relies on an alteration of the process and then weighting samples using the likelihood-ratio between the original and the altered process.

4.2 MOMENT CONSTRAINTS

The time-evolution of $E(f(X_t))$ for some function f is given by (2.19). The integration of (2.19) with such functions f is well-known in the context of moment approximations of MPM. For most models the arising ODE system is infinitely large, because the time-derivative of low order moments usually depends on the values of higher order moments. To close this system, *moment closures*, i.e. ad-hoc approximations of higher order moments are applied [SSG15]. The main drawback of this kind of analysis is that it is not known whether the chosen closure gives an accurate approximation for the case at hand. Here, such approximations are not necessary, since we will apply the moment dynamics in the context of stochastic sampling instead of trying to integrate (2.19).

Apart from integration strategies, setting (2.19) to zero has been used as a constraint for parameter estimation at steady-state [BBW18] and bounding moments at steady-state [DB18a; Ghu+17; Kun+19a]. The extension of the latter has recently lead to the adaption of these constraints to a transient setting [DB18b; SH19]. These two transient constraint variants are analogously derived by multiplying (2.19) by a time-dependent, differentiable weighting function $w(t)$ and integrating:

Multiplying with $w(t)$ and integrating on $[t_0, T]$ yields [DB18b; SH19]

$$w(T)E\left(f(X_T)\right) - w(t_0)E\left(f(X_{t_0})\right) - \int_{t_0}^{T} \frac{dw(t)}{dt} E\left(f(X_t)\right) dt$$

$$= \sum_{j=1}^{n_R} \int_{t_0}^{T} w(t)E\left(\left(f(X_t + v_j) - f(X_t)\right)\alpha_j(X_t)\right) dt \tag{4.1}$$

In the context of computing moment bounds via semi-definite programming the polynomial $w(t) = t^s$ [SH19] and the exponential $w(t) = e^{\lambda(T-t)}$ [DB18b] have been proposed. While both choices proved to be effective in different case studies, relying solely on the latter choice, i.e.

$$w(t) = e^{\lambda(T-t)}$$

was sufficient.

By expanding the rate functions and f in (4.1) and substituting the exponential weight function we can re-write (4.1) as

$$0 = E\left(f(X_T)\right) - e^{\lambda T}E\left(f(X_{t_0})\right) + \sum_k c_k \int_{t_0}^{T} e^{\lambda(T-t)} E\left(X_t^{m_k}\right) dt \tag{4.2}$$

with coefficients c_k and vectors m_k defined accordingly. Assuming the moments remain finite on $[0, T]$, we can define the random variable

$$Z = f(X_T) - e^{\lambda T}f(X_{t_0}) + \sum_k c_k \int_{t_0}^{T} e^{\lambda(T-t)} X_t^{m_k} dt \tag{4.3}$$

with $E(Z) = 0$.

Note, that a realization of Z depends on the whole trajectory $\tau = x_0 t_1 x_1 t_1 \dots t_n x_n$ over $[t_0, T]$. Thus, for the integral terms in (4.3) we have to compute sums

$$\frac{1}{\lambda} \sum_{i=1}^{n} \left(e^{\lambda(T-t_{i+1})} - e^{\lambda(T-t_i)}\right) x_i^{m_k}, \tag{4.4}$$

over a given trajectory. This accumulation is best done during the simulation to avoid storing the whole trajectory. Still, the cost of a simulation run increases. For the method to be efficient, the variance reduction (Section 4.3) needs to overcompensate for this increased cost of a simulation run.

EXAMPLE   For Model 1 the moment equation for $f(x) = x$ becomes

$$\frac{d}{dt} E(X_t) = \gamma - \delta E(X_t) .$$

The corresponding constraint (4.2) with $\lambda = 0$ gives

$$0 = E(X_T) - E(X_0) - \gamma T + \delta \int_0^T E(X_t) \, dt .$$

In this instance the constraint leads to an explicit function of the moment over time. If $X_0 = 0$ w.p. 1, then (4.2) becomes

$$E(X_T) = \frac{\gamma}{\delta} \left(1 - e^{-\delta T}\right) \tag{4.5}$$

when choosing $\lambda = -\delta$.   ◇

4.3   CONTROL VARIATES

Now, we are interested in the estimation of some quantity $E(V)$ by stochastic simulation. Let $V_1, \ldots, V_n$ be independent samples of V. Then the sample mean $\hat{V}_n = \frac{1}{n} \sum_{i=1}^n V_k$ is an estimate of $E(V)$. By the central limit theorem

$$\sqrt{n}\hat{V}_n \xrightarrow{d} N(E(V), \sigma_V^2) .$$

Now suppose, we know of a random variable Z with $0 = E(Z)$. The variable Z is called a *control variate (CV)*. If a control variate Z is correlated with V, we can use it to reduce the variance of $\hat{V}_n$ [GY05; Nel90; Sze03; Wil84]. For example, consider we are running a set of simulations and consider a single constraint. If the estimated value of this constraint is larger than zero and we estimate a positive correlation between the constraint Z and V, we would, intuitively, like to decrease our estimate $\hat{V}_n$ accordingly. This results in an estimation of the mean of the random variable

$$Y_\beta = V - \beta Z$$

instead of V. The variance

$$\sigma_{Y_\beta}^2 = \sigma_V^2 - 2\beta \text{Cov}(V, Z) + \beta^2 \sigma_Z^2 .$$

The optimal choice β can be computed by considering the minimum of $\sigma_{Y_\beta}^2$. Then

$$\beta^* = \text{Cov}(V, Z)/\sigma_Z^2.$$

Therefore $\sigma_{Y_{\beta^*}} = \sigma_Z^2(1 - \rho_{VZ}^2)$, where ρ_{VZ} is the correlation of Z and V.

If we have multiple control variates, we can proceed in a similar fashion. Now, let Z denote a vector of d control variates and let

$$\Sigma = \begin{bmatrix} \Sigma_Z & \Sigma_{VZ} \\ \Sigma_{ZV} & \sigma_V^2 \end{bmatrix}$$

be the covariance matrix of (Z, V). As above, we estimate the mean of $Y_\beta = V - \beta^\top Z$. The ideal choice of β is the result of an ordinary least squares regression between V and Z_i, $i = 1, \dots, n$. Specifically,

$$\beta^* = \Sigma_Z^{-1}\Sigma_{ZV}.$$

Then, asymptotically the variance of this estimator is [Sze03],

$$\sigma_{\hat{Y}_{\beta^*}}^2 = (1 - R_{ZV}^2)\sigma_{\hat{V}}^2, \quad R_{ZV}^2 = \Sigma_{ZV}\Sigma_Z^{-1}\Sigma_{ZV}/\sigma_V^2. \qquad (4.6)$$

In practice, however, β^* is unknown and needs to be replaced by an estimate $\hat{\beta}$. This leads to an increase in the estimator's variance. Under the assumption of Z and V having a multivariate normal distribution [Che78; LMW82], the variance of the estimator is $\hat{Y}_{\hat{\beta}} = \hat{V} - \hat{\beta}^\top \hat{Z}$

$$\sigma_{\hat{Y}_{\hat{\beta}}}^2 = \frac{n-2}{n-2-d}(1 - R_{ZV}^2)\sigma_{\hat{V}}^2. \qquad (4.7)$$

Clearly, a control variate is "good" if it is highly correlated with V. The constraint in (4.5) is an example of the extreme case. When we use this constraint as a control variate for the estimation of the mean at some time point t, it has a correlation of ± 1 since it describes the mean at that time precisely. Therefore the variance is reduced to zero. We thus aim to pick control variates that are highly correlated with V.

EXAMPLE   Consider, for example, the above case of the birth-death process. If we choose (4.5) as a constraint, it would always yield the exact difference of the exact mean to the sample mean and therefore have a perfect correlation. Clearly, $\hat{\beta}$ reduces to 1 and $\hat{Y}_1 = E(X_t)$. ◇
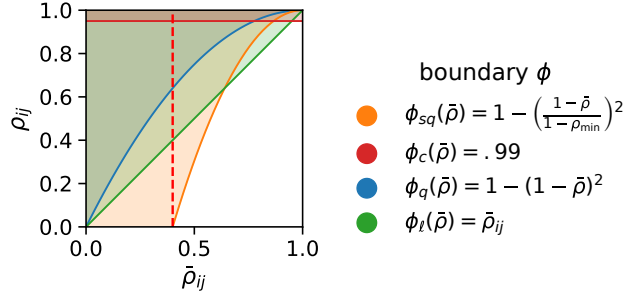
Figure 4.1: Different decision functions used in the redundant control variate removal. The weaker of any two control variates is removed if the pair $(\bar{\rho}_{ij}, \rho_{ij})$ belongs to the shaded area of the considered function. The vertical dashed line indicates $\rho_{\min}$.

4.4 MOMENT-BASED VARIANCE REDUCTION

We propose an adaptive estimation algorithm (Algorithm 2) that starts out with an initial set of control variates and periodically removes potentially inefficient variates. The "accumulator set" A represents the time-integral terms (4.4). The size of A has the most significant impact on the overall speed of the algorithm since it represents the only factor incurring a direct cost increase in the SSA itself (line 4).

The algorithm consists of a main loop which performs n simulation runs (line 3). Between each run the mean and covariance estimates of $[Z, V]$ are updated (line 5). Every $d < n$ iterations, the control variates are checked for *efficiency* and *redundancy* (lines 6–10).

Checking both conditions is based on the correlation ρ_{ij} between the i-th and j-th control variate and the correlation ρ_{iv} of a control variate i to V. The first condition is a simple lower threshold $\rho_{\min}$ for a correlation ρ_{iv}. This condition aims to remove those variates from the control variate set that are only weakly correlated to V (line 8). The rationale is that, if variate i has a low correlation with the variable of interest V, its computation may
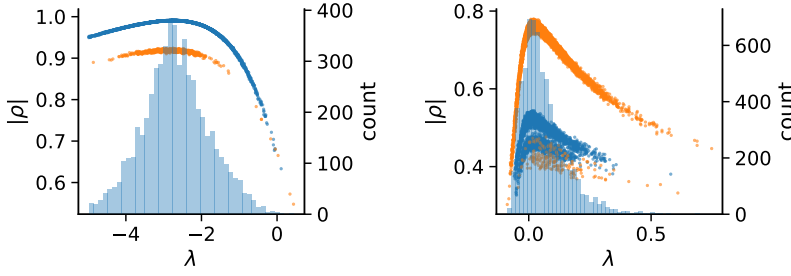
Figure 4.2: The absolute correlation of different constraints to V arising from different choices of λ. The blue dots represent constraints based on first order moments, while the orange refers to control variates derived from second order moments. In both cases 10,000 samples were used with 30 initial samples for λ from $N(0,1)$ and $k_{min} = 2$. A quadratic decision bound was used for the redundancy removal. Furthermore, a histogram of control variates selected by Algorithm 2 is given. In (a) $E\left(X_2^A\right)$ in the dimerization model was estimated. In (b) $E\left(X_{50}^X\right)$ in the processive modification model was estimated.

not be worth the costs. Here, we propose to set ρ_{min} heuristically as

$$\rho_{min} = \min\left(0.1, \frac{\max_i \rho_{iv}}{k_{min}}\right),$$

where $k_{min} > 1$ is an algorithm parameter.

The second condition aims to remove redundant conditions. This is not only beneficial for the efficiency of the estimator, but also necessary for the matrix inversion (4.6) because perfectly and highly correlated constraints will make the covariance matrix estimate $\hat{\Sigma}_Z$ (quasi-) singular. For all considered criteria we iterate over all tuples $(i, j) \in \{1, \ldots, k\}^2$, $i \neq j$, removing the weaker of the two, i.e. $\arg\min_{k \in \{i,j\}} \rho_{kv}$, if the two control variates are considered redundant (line 9).

There are many ways to define such a redundancy criterion. Here, we focus on criteria that are defined in terms of the average correlation

$$\bar{\rho}_{ij} = (\rho_{iv} + \rho_{jv})/2.$$

For two variates i and j we then check if their mutual correlation ρ_{ij} exceeds a some function ϕ of $\bar\rho_{ij}$, i.e. we check the inequality

$$\phi(\bar\rho_{ij}) \leqslant \rho_{ij}.$$

If this inequality holds, constraint $\arg\min_{k\in\{i,j\}} \rho_{kv}$ is removed. Naturally, there are many possible choices for the above decision boundary ϕ (cf. Figure 4.1).

The simplest choice is to ignore $\bar\rho_{ij}$ and just fix a constant close to 1 as a threshold, e.g.

$$\phi_c(\bar\rho_{ij}) = .99.$$

While this often leads to the strongest variance reduction and avoids numerical issues in the control variate computation, it turns out that the computational overhead is not as well-compensated as by other choices of ϕ (see Section 4.5).

Another option is to fix a simple linear function, i.e.

$$\phi_\ell(\bar\rho_{ij}) = \bar\rho_{ij}.$$

For this choice the intuition is, that one of two constraints is removed if their mutual correlation exceeds their average correlation with V.

Here, we also assess two quadratic choices for ϕ. The first choice of

$$\phi_q(\bar\rho) = 1 - (1 - \bar\rho)^2$$

is more tolerant than the linear function and more strict than a threshold function, except for highly correlated control variates. Another variant of ϕ is given by including the lower bound $\rho_{\min}$ and scaling the quadratic function accordingly:

$$\phi_{sq}(\bar\rho) = 1 - ((1 - \bar\rho)/(1 - \rho_{\min}))^2.$$

The different choices of ϕ considered here are plotted in Figure 4.1.

Now, we discuss the choice of the initial control variates. We identify control variate k by a tuple (m_k, λ_k) of a moment vector m_k and a time-weighting parameter λ_k. That is, we use $w(t) = e^{\lambda_k(T-t)}$ and $f(x) = x^{m_k}$ in (4.1). For a given set of parameters L, we use all moments up to some fixed order $n_{\max}$ (line 2). The

---

Algorithm 2: Estimate the mean of species i at time T

---

input : $n, d, n_{max}, n_\lambda, k_{min}$
output: An estimate using linear control variates

1 $L = \{\lambda_i \sim \pi_\lambda \mid 1 \leqslant i < n_\lambda\} \cup \{0\}$;
2 $P \leftarrow \{(m, \lambda) \mid 1 \leqslant |m| \leqslant n_{max}, \lambda \in L\}$;
3 **for** $i = 1, \ldots, n$ **do**
4 $\quad$ $\tau \leftarrow SSA(\pi_0, T, A)$;
5 $\quad$ compute constraint values using A and update $\hat{\Sigma}$ and $\hat{V}_i$;
6 $\quad$ **if** $i \mod d = 0$ **then**
7 $\quad\quad$ $\rho_{min} \leftarrow \min(0.1, \max_i \rho_{iv}/k_{min})$;
8 $\quad\quad$ $P \leftarrow P \setminus \{(m_k, \lambda_k) \mid \rho_{kv} < \rho_{min}\}$;
9 $\quad\quad$ $P \leftarrow P \setminus \{(m_k, \lambda_k) \mid \exists i, j. i \neq j, \phi(\bar{\rho}_{ij}) < \rho_{ij}, k = \arg\min_{k \in \{i,j\}} \rho_{kv}\}$;
10 $\quad$ remove unneeded accumulators from A;

11 **return** $\hat{V}_n - (\hat{\Sigma}_Z^{-1} \hat{\Sigma}_{ZV})^\top \hat{Z}_n$;

---

ideal set of parameters L is generally not known. For certain choices the correlation of the control variates and the variable of interest is higher then for others. To illustrate this, consider the above example of the birth-death process. Choosing $\lambda = -\delta$ leads to a control variate that has a correlation of ± 1 with V. Therefore, the ideal choice of initial values for would be $L = \{-\delta\}$. This, however, is generally not known. Therefore, we sample a set of λ's from some fixed distribution π_λ (line 1).

4.5 CASE STUDIES

We first define a criterion of *efficiency* in order to estimate whether the reduction in variance is worth the increased cost. A natural baseline of a variance reduction is, that it is more efficient to pay for the overhead of the reduction than to generate more samples to achieve a similar reduction of variance. Let σ_Y^2 be the variance of Y. The *efficiency* of the method is the ratio of the necessary

cost to achieve a similar reduction with the CV estimate Y_{CV} compared to the standard estimate Y [L'E94], i.e.

$$E = \frac{c_0 \sigma_Y^2}{c_1 \sigma_{Y_{CV}}^2} .$$

(4.8)

That ratio c_0/c_1 depends heavily on both the specific implementation and the technical setup. The cost increase is mainly due to the computation of the integrals in (4.2). But the repeated checking of control variates for efficiency also increases the cost. The accumulation over the trajectory directly increases the cost of a single simulation which is the critical part of the estimation. To estimate the base-line cost c_0, 2000 estimations were performed without considering any control variates.

The simulation is implemented in the Rust programming language[1]. The model description is parsed from a high level specification. Rate functions are compiled to stack programs for fast evaluation. Code is made available online [Bac19].

We consider four non-trivial case studies. Three models exhibit complex multi-modal behaviour. We now describe the models and the estimated quantities in detail.

This model appeared earlier as Model 2 on page 27.

The first model is a simple dimerization on a countably infinite state-space.

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

$$\varnothing \xrightarrow{10} M, \quad 2M \xrightarrow{0.1} D$$
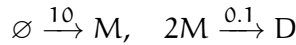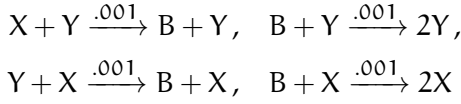
with initial condition $X_0^M = 0$.

Despite the models simplicity, the moment equations are not closed for this system due to the second reaction which is non-linear. Therefore a direct analysis of the expected value would require a closure. For this model we will estimate $E\left(X_2^M\right)$.

The following two models are bimodal, i.e. they each posses two stable regimes among which they can switch stochastically. For both models we choose the initial conditions such that the process will move towards either attracting region with equal probability.
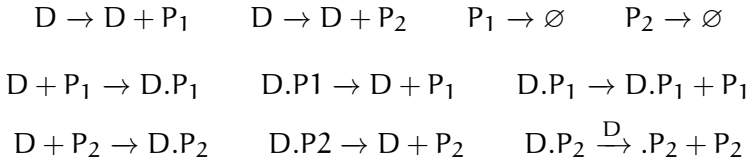
---

1 https://www.rust-lang.org

Model 5 (Distributive Modification). *The distributive modification model was introduced in Cardelli and Csikász-Nagy [CCN12]. It consists of the reactions*

$$X + Y \xrightarrow{.001} B + Y, \quad B + Y \xrightarrow{.001} 2Y,$$
$$Y + X \xrightarrow{.001} B + X, \quad B + X \xrightarrow{.001} 2X$$

with initial conditions $X_0^X = X_0^Y = X_0^B = 100$.

Model 6 (Exclusive Switch). *The exclusive switch model consists of 5 species, 3 of which are typically binary (activity states of the genes) [Loi+07].*

$$D \to D + P_1 \quad D \to D + P_2 \quad P_1 \to \varnothing \quad P_2 \to \varnothing$$

$$D + P_1 \to D.P_1 \quad D.P1 \to D + P_1 \quad D.P_1 \to D.P_1 + P_1$$

$$D + P_2 \to D.P_2 \quad D.P2 \to D + P_2 \quad D.P_2 \xrightarrow{D} .P_2 + P_2$$

with initial conditions $X_0^D = 1$ and $X_0^{D.P1} = X_0^{D.P_2} = X_0^{P_1} = X_0^{P_2} = 0$.

We evaluate the influence of algorithm parameters, choices of distributions to sample λ from, and the influence of the sample size on the efficiency of the proposed method. Note that the implementation does not simplify the constraint representations or the state space according to stoichiometric invariants or limited state spaces. Model 5, for example has the invariant $X_t^X + X_t^Y + X_t^B = $ const., $\forall t \geqslant 0$, which could be used to reduce the state-space dimensionality to two. In Model 6 the invariant $\forall t \geqslant 0.X_t^D, X_t^{D.P_1}, X_t^{d.P_2} \in \{0,1\}$ could be used to optimize the algorithm by eliminating redundant moments, e.g. $E((X^D)^2) = E(X^D)$. Such an optimization would further increase the efficiency of the algorithm.

We first turn to the choice of the λ sampling distribution. Here we consider two choices:

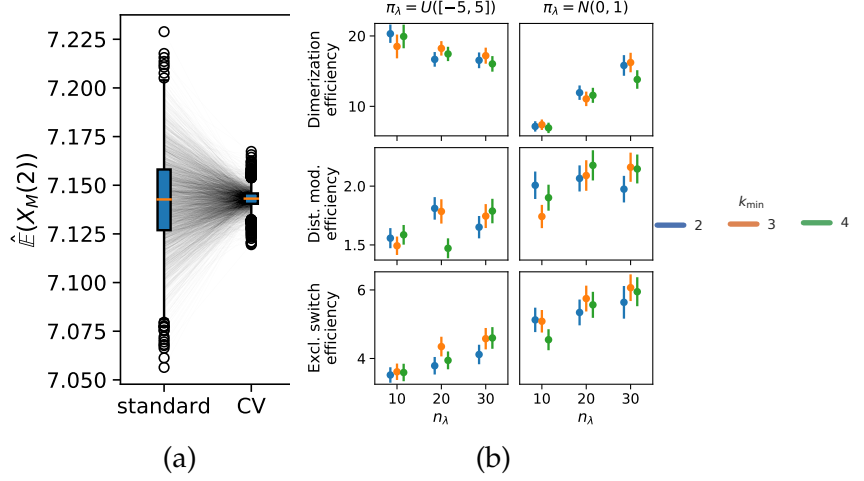1. a standard normal distribution $N(0,1)$,

2. a uniform distribution on $[-5,5]$.

Figure 4.3: (a) The effect of including CVs on the mean estimates $\hat{E}(X_2^M)$ in the dimerization case study. Parameters were $\pi_\lambda = N(0, 1)$, $n_\lambda = 30$, $k_{min} = 4$, $\phi(\bar{\rho}) = 1 - (1 - \bar{\rho})^2$. (b) The efficiencies for different valuations of n_λ and k_{min} and choices of π_λ. The sample size was $n = 10,000$ in all cases with $d = 100$. The bars give the bootstrapped (1000 iterations) standard deviations.

We deterministically include $\lambda = 0$ in the constraint set, as this parameter corresponds to a uniform weighting function. We performed estimations on the case studies using different valuations of the algorithm parameters of the minimum threshold k_{min} and the number of λ-samples n_λ. We used samples size $n = 10,000$ and checked the control variates every $d = 100$ iterations for the defined criteria. For each valuation 1000 estimations were performed. In Figure 4.3b, we summarize the efficiencies for the arising parameter combinations on the three case studies. Most strikingly, we can note that the efficiency was consistently larger than one in all cases. Generally, the normal sampling distribution out-performed the alternative uniform distribution, except in case of the dimerization. The reason for this becomes apparent, when examining Figure 4.2: In case of the dimerization model the most efficient constraints are found for $\lambda \approx -3$, while in case of the distributive modification they are located just above 0 (we observe a similar pattern for the exclusive switch case study). Therefore the sampling of efficient λ values is more likely using a

uniform distribution for the dimerization case study, than it is for the others. Given that larger absolute values for λ seem unreasonable due their exponential influence on the weighting function and the problem of fixing a suitable interval for a uniform sampling scheme, the choice of a standard normal distribution for π_λ seems superior.
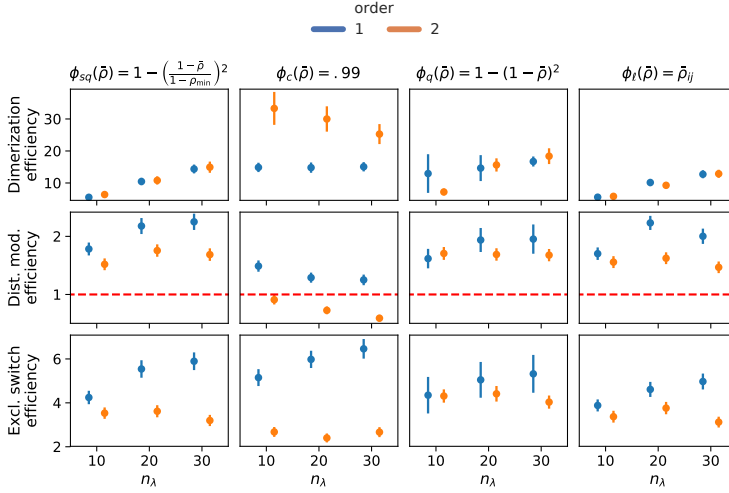

Figure 4.4: The efficiency for different redundancy policies ϕ and maximal moment orders n_{max}. The sample size was $n = 10{,}000$ in all cases with $d = 100$. Furthermore, $k_{min} = 3$, $\pi_\lambda = N(0,1)$, and $n_{max} = 1$. The bars give the bootstrapped (1000 iterations) standard deviations.

In Figure 4.4 we compare efficiencies for different maximum orders of constraints n_{max}. This comparison is performed for different choices of the redundancy rule and initial λ sample sizes n_λ. Again, for each parameter valuation 1000 estimations were performed. With respect to the maximum constraints order n_{max} we see a clear tendency, that the inclusion of second order constraints lessens the efficiency of the method. In case of a constant redundancy threshold it even dips below break-even for the distributive modification case study. This is not surprising, since the inclusion of second order moments increases the number of initial constraints quadratically and the incurred cost, especially of the first iterations, lessens efficiency.
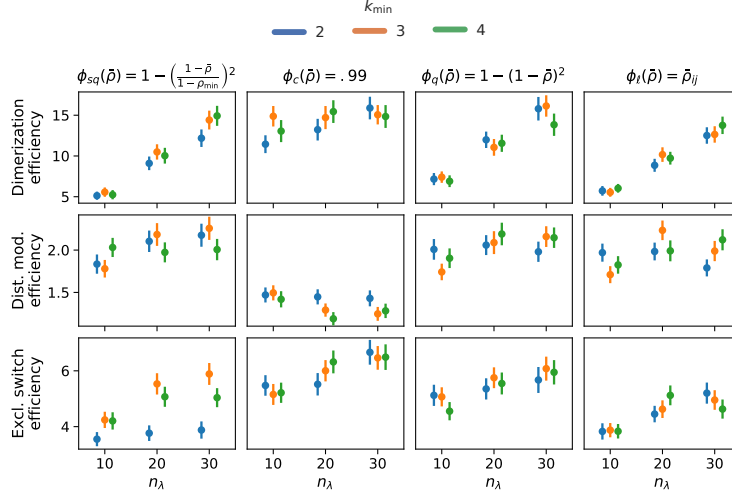
Figure 4.5: The empirical efficiencies for different n_λ and k_{min}. On the considered case studies. The sample size was $n = 10,000$ in all cases with $d = 100$. 1000 estimations were performed for each case. The bars give the bootstrapped (1000 iterations) standard deviations. The break-even $E = 1$ is marked by the dotted red line.

Figure 4.6 depicts the trade-off between the variance reduction σ_0^2/σ_1^2 versus the cost ratio c_0/c_1. Comparing the redundancy criterion based on a constant threshold ϕ_c to the others, we observe both a larger variance reduction and an increased cost. This is due to the fact, that more control variates are included throughout the simulations (Table 3, Table 4). Depending on the sample distribution π_λ and the case study, this permissive strategy may pay off. In case of the dimerization, for example, it pays off, while in case of the distributive modification it leads to a lower efficiency ratio. In the latter case the model is more complex, and therefore the set of initial control variates is larger. With a more permissive redundancy strategy, more control variates are kept (ca. 10 when using ϕ_c vs. ca. 2–3 for the others). The other redundancy boundaries move the results further in the direction of less variance reduction while keeping the cost increase low. On the opposite end is the linear ϕ_ℓ. The quadratic versions ϕ_q and ϕ_{sq} can be found in the middle of this spectrum.
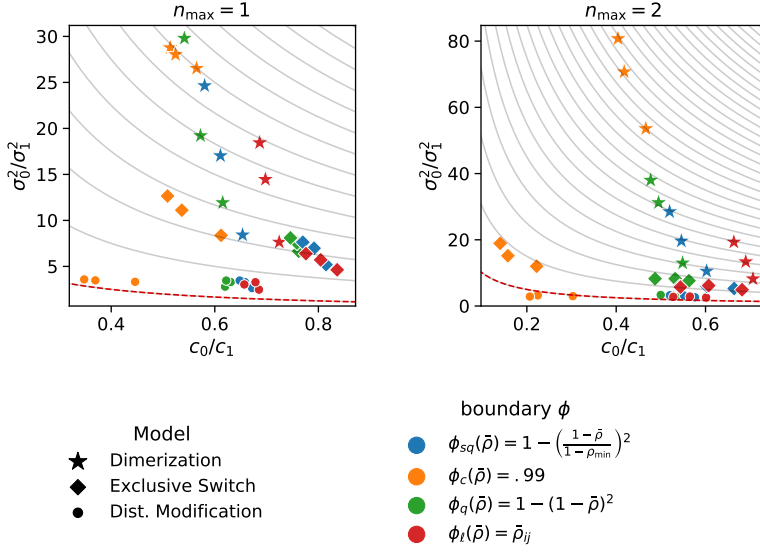
Figure 4.6: A visualisation of the trade-off between variance reduction σ_0^2/σ_1^2 and cost ratio c_0/c_1. Isolines for efficiencies are given in grey. The break-even is marked by the dashed red line. Markers of the same kind differ in n_λ and shift with increasing value upwards in variance reduction and lower in c_0/c_1, i.e. the shift is to the left and upwards with increasing n_λ. The sample size was $n = 10,000$ in all cases with $d = 100$. Furthermore, $k_{min} = 3$ and $\pi_\lambda = N(0,1)$.

We also observe, that an increase of n_λ is particularly beneficial, if the sampling distribution π_λ does not capture the parameter region of the highest correlations well. This can be seen for the Dimerization case study, where the variance reduction increases strongly with increasing sample size (Figure 4.5, Table 3, Table 4). Since $\pi_\lambda = N(0,1)$, more samples are needed to sample efficient λ-values (cf. Figure 4.2).

In Figure 4.5 we give detailed information on the influence of algorithm parameters k_{min}, the number of initial λ values, and different redundancy rules. The λ sampling distribution π_λ is a standard normal.

Finally, we discuss the effect of the sample size n on the efficiency E. In Figure 4.7 we give both the efficiencies and the slowdown for different sample sizes. As a redundancy rule we used the unscaled quadratic function, 30 initial values of λ, and

$k_{min} = 3$. With increasing sample size, the efficiency usually approaches an upper limit. This is due to the fact that most control variates are dropped early on and the control variates often remain the same for the rest of the simulations. If we assume there are no helpful control variates in the initial set and all would be removed at iteration 100, the efficiency would approach 1 with $n \to \infty$.
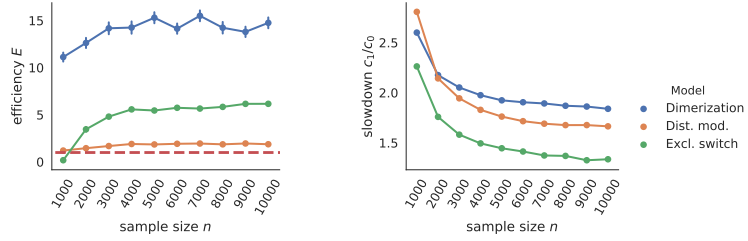
(a) Efficiency v. sample size   (b) Slowdown v. sample size

Figure 4.7: The effect of sample size on the efficiency E and slowdown in the different case studies. The break-even $E = 1$ is marked by the dashed red line. The cost increase due to the variance reduction over different sample sizes.

4.6 CONCLUSION

In this work we have shown that known constraints on the moment dynamics can be successfully leveraged in simulation-based estimation of expected values. The empirical results indicate that the supplementing a standard SSA estimation with moment information can drastically reduce the estimators' variance. This reduction is paid for by accumulating information on the trajectory during simulation. However, the reduction is able to compensate for this increase. This means that for fixed costs, using estimates with control variates is more beneficial than using estimates without control variates.

While a variety of algorithmic variants was evaluated, many aspects remain subject to further study. In particular different choices of f and w in (4.1) may improve efficiency further. These choices become particularly interesting when moving from the

estimation of simple first order moments to more complex queries such as behavioural probabilities of trajectories. In such cases, one might even attempt to find efficient control variate functions using machine learning methods.

Another open question regarding this work is its performance when multiple quantities instead of a single quantity are to be estimated. In such a case, constraints would be particularly beneficial, if they lead to improvements as many estimation targets as possible.

Furthermore the identification of the best weighting parameters λ could be done in a more adaptive fashion. The presented scheme of a sampling from π_λ could be extended into a Bayesian-like procedure, wherein the values for λ are periodically re-sampled from a distribution that is adjusted according to the best-performing constraints up to that point.

Part III

AGGREGATION & REFINEMENT

We present a state-space lumping scheme that aggregates states in a grid structure. Approximations based on this lumping are used to iteratively refine relevant and truncate irrelevant parts of the state-space. This way, the algorithm learns a well-justified finite-state projection for different scenarios.

STATE-SPACE AGGREGATION

In this part, we propose a method to identify a truncation that optimizes the trade-off between the size of the considered state-space and the approximation error due to the finite state projection (FSP). To this end, we start with a very coarse-grained model abstraction that we refine iteratively. The coarse-grained model is based on an grid-shaped aggregation (i.e. lumping) scheme that identifies a set of macro-states. These macro-states can be used to compute an interim model solution that guides the refinement in the next step. We perform refinements until the approximation arrives at the resolution of the original model (i.e. each macro-state has only one constituent) such that the aggregation introduces no approximation error.

5.1 MACRO-STATES

A macro-state is a collection of micro-states (or simply states) treated as one state in the aggregated model, which can be seen as an abstraction of the original model. The aggregation scheme defines a partitioning of the state-space. We choose a scheme based on a grid structure. That is, each macro-state is a hypercube in $\mathbb{Z}_{\geqslant 0}^{n_s}$.

Hence, each macro-state $\bar{x}_i(\ell^{(i)}, u^{(i)})$ (denoted by $\bar{x}_i$ for notational ease) can be identified using two vectors $\ell^{(i)}$ and $u^{(i)}$. The vector $\ell^{(i)}$ gives the corner closest to the origin, while $u^{(i)}$ gives the corner farthest from the origin. Formally,

$$\bar{x}_i = \bar{x}_i(\ell^{(i)}, u^{(i)}) = \{x \in \mathbb{N}^{n_s} \mid \ell^{(i)} \leqslant x \leqslant u^{(i)}\}, \qquad (5.1)$$

where $'\leqslant'$ denotes element-wise comparison.

In order to solve the aggregated model, we need to define transition rates between macro-states. Therefore, we assume that, given that the system is in a particular macro-state, all constituent states are equally likely (uniformity assumption). This assump-

tion is the reason why the aggregated model provides only a coarse-grained approximation.

The uniformity assumption is a modeling choice yielding significant advantages. Firstly, it eases the computation of the rates between macro-states and, therefore, makes a fast solution of the aggregated model possible. Secondly, even though it induces an approximation error, it provides suitable guidance as uniformity assumption spreads out the probability mass conservatively. Hence, it becomes less likely that regions of interest are disregard. Lastly, the uniformity assumption is theoretically well-founded, as it stems from the maximum entropy principle: In the absence of concrete knowledge about the probability distribution inside a macro-state, we assume the distribution with the highest uncertainty, i.e., the uniform distribution.

5.2   CONSTRUCTION

The grid structure makes the computation of transition rates between macro-states particularly convenient and computationally simple. Mass-action reaction rates can be given in a closed-form, due to the Faulhaber formulae [Knu93] and more complicated rate functions such as Hill-functions can often be handled as well by taking appropriate integrals (see Section 7.5.2.2).

Suppose, we are interested in the transition rate from macro-state $\bar{x}_i$ to macro-state $\bar{x}_k$ according to reaction j. Using the uniformity assumption, this is simply the mean rate of the states in $\bar{x}_i$ that go to $\bar{x}_k$ using j. However, only a small subset of constituents in $\bar{x}_i$ are actually relevant for this transition. Hence, we identify the subset of states of $\bar{x}_i$ that lie at the border to $\bar{x}_k$ and in such a way that applying reaction j shifts them to a state in $\bar{x}_k$. Then, we sum up the corresponding rates of these states. Lastly, we normalize according to the number of states inside of $\bar{x}_i$.

It is easy to see that the relevant set of border states is itself an interval-defined macro-state $\bar{x}_{i \xrightarrow{j} k}$. To compute this macro-state we can simply shift $\bar{x}_i$ by v_j, take the intersection with $\bar{x}_k$ and project this set back. Formally,

$$\bar{x}_{i \xrightarrow{j} k} = ((\bar{x}_i + v_j) \cap \bar{x}_k) - v_j \,, \tag{5.2}$$

where the additions are applied element-wise to all states making up the macro-states. For ease of notation, we also define a general exit state

$$\bar{x}_{i \xrightarrow{j}} = ((\bar{x}_i + v_j) \setminus \bar{x}_i) - v_j. \tag{5.3}$$

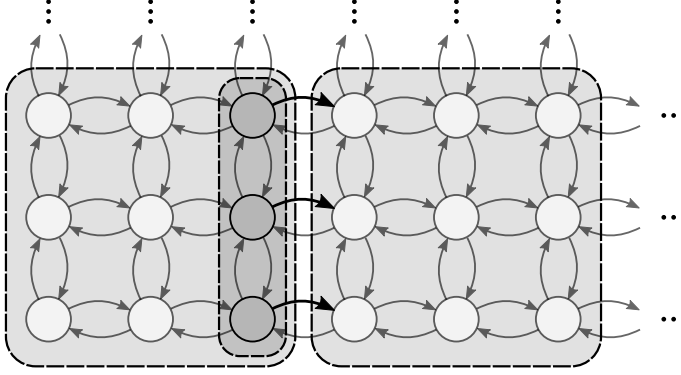This state captures all micro-states inside $\bar{x}_i$ that can leave the state via reaction j.

Figure 5.1: Two macro-states and a transition state from the left to the right.

This uniformity assumption gives rise to the following Q-matrix of the aggregated model:

$$\bar{Q}_{\bar{x}_i, \bar{x}_k} = \begin{cases} \sum_{j=1}^{n_R} \bar{\alpha}_j \left(\bar{x}_{i \xrightarrow{j} k} \right) / |\bar{x}_i| , & \text{if } \bar{x}_i \neq \bar{x}_k \\ -\sum_{j=1}^{n_R} \bar{\alpha}_j \left(\bar{x}_{i \xrightarrow{j}} \right) / |\bar{x}_i| , & \text{otherwise} \end{cases} \tag{5.4}$$

where

$$\bar{\alpha}_j (\bar{x}) = \sum_{x \in \bar{x}} \alpha_j (x). \tag{5.5}$$

is the sum of all rates belonging to reaction j in $\bar{x}$.

Under the assumption of polynomial rates, as is the case for mass-action systems, we can compute the sum of rates over this transition set efficiently using Faulhaber's formula. As an example consider the following mass-action reaction $2X \xrightarrow{c} \varnothing$.
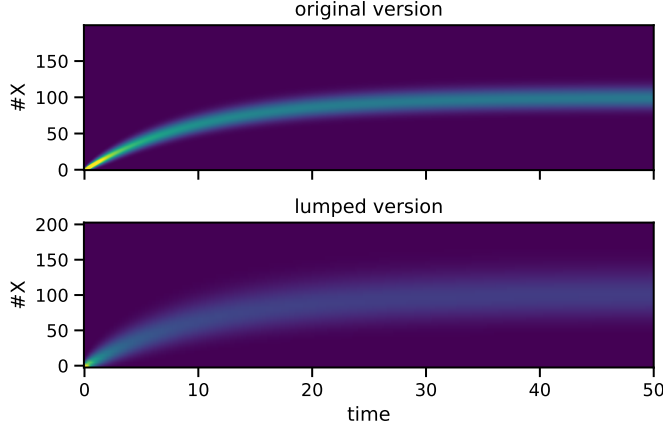
Figure 5.2: A lumping approximation of Model 1 on the state-space truncation to $[0, 200]$ on $t \in [0, 50]$. On the left-hand side solutions of a regular truncation approximation and a lumped truncation (macro-state size is 5) are given.

For macro-state $\bar{x} = \{0, \dots, n\}$ we can compute the corresponding lumped transition rate

$$\bar{\alpha}(\bar{x}) = \frac{c}{2} \sum_{i=1}^{n} i(i-1) = \frac{c}{2} \left(\sum_{i=1}^{n} i^2 - \sum_{i=1}^{n} i \right)$$
$$= \frac{c}{2} \left(\frac{2n^3 + 3n^2 + n}{6} - \frac{n^2 + n}{2} \right)$$

eliminating the explicit summation in the lumped propensity function.

Interestingly, the lumped distribution tends to be less concentrated. This is due to the assumption of a uniform distribution inside macro-states. This effect is illustrated by the example of a birth-death process in Figure 5.2 and Figure 5.3. Due to this effect, an iterative refinement typically keeps an over-approximation in terms of state-space area. This is a desirable feature since relevant regions are less likely to be pruned due to lumping approximations.
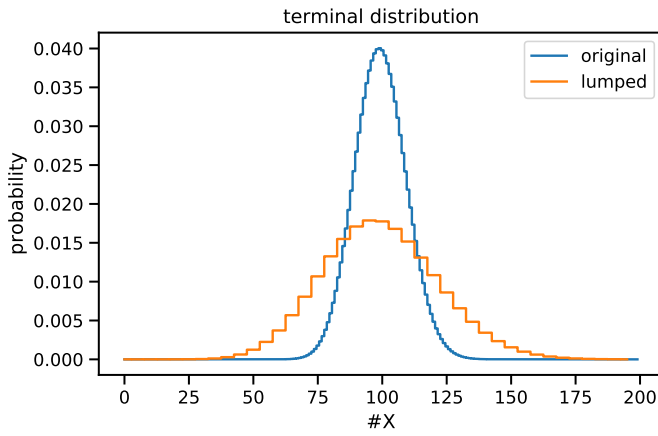
Figure 5.3: The terminal distributions $\Pr(X_{50} = x_i)$ of a lumped approximation and a truncation at original granularity are given.

6

TRUNCATIONS FOR STATIONARY DISTRIBUTIONS

An important part of the analysis of such models concerns their long-run behavior. Given an ergodic underlying Markov chain, the chain's stationary distribution characterizes this behavior. For some special model classes, such as zero-deficiency networks [AK11], analytical solutions for the stationary distribution are known. However, most models require numerical approaches, often based on some form of approximation to guarantee tractability. Those approaches can be based on stochastic simulation [Gil77] (which for steady-state analysis tends to be slow and inaccurate) or moment-bounds via mathematical programming [Kun+19a]. Here, we draw on numerical approaches based on state-space truncation, which represent a viable option to approximate stationary distributions [Kun+21a]. Truncation-based approaches have the benefit of describing the complete dynamics within a finite subset of the typically very large or infinite state-space. As such, they enable the approximation of complex distributions that are not well-described by low-order moments.

The main step in the computation of such an approximation is the identification of a suitable truncation, i.e. a subset of the state-space encompassing most of the stationary probability mass. Existing methods typically rely on Foster-Lyapunov drift conditions to define such subsets [Day+11]. While these truncations come with bounds on the contained stationary probability mass, they typically are far larger than necessary. The truncation is usually strongly constrained by the form of the chosen Lyapunov function [GBK14; Day+11]. Optimizing over possible functions to identify efficient truncations is technically challenging and, to our knowledge, has not been demonstrated for general reaction networks [MAK14].

In this work, we address the identification of suitable truncations by using an aggregation-refinement scheme. Initially, a

Lyapunov analysis yields a set containing at least $1 - \epsilon$ of the stationary probability mass. On this subset of the state-space, we apply an aggregation scheme that groups together states in hypercube macro-states. Throughout each of these macro-states, we assume a uniform distribution among its constituent micro-states. This allows us to roughly analyze large portions of the state-space with exponentially fewer variables. We then iteratively truncate and refine the approximation based on the stationary distribution of this aggregated Markov chain. We keep only the most relevant macro-states and continue this scheme until the macro-states contain a single original state. In this way, we arrive at an effective truncation to compute an approximation of the stationary distribution.

We investigate the approximation results on case studies with known stationary distributions and complex models with intricate stationary distributions. We evaluate the truncation quality by assessing the stationary probability mass captured. To this end, we use analytical solutions and bounds given by a Lyapunov analysis. Further, we explore the control of the truncation size through the truncation parameter. Finally, we demonstrate the method on the p53 oscillator model exhibiting a complex stationary distribution.

6.1 RELATED WORK

ANALYTICAL SOLUTIONS   For some specific models, analytical solutions for the stationary distribution have been found [MHK14; Kur+18]. For the class of zero-deficiency networks, the stationary distribution is known to have a Poisson product form [ACK10]. Monomolecular reaction networks can be solved explicitly, as well [JH07].

TRUNCATION-BASED ANALYSIS   The analysis of countably infinite-sized state-spaces is often handled by pre-defined truncations [KNP11]. Sophisticated state-space truncations for the (unconditioned) forward analysis have been developed that give lower bounds. They typically provide a trade-off between computational load and tightness of the bound [MK06; LMW11; And+11; HMW09; Mik+13]. Such methods cannot be directly

applied to the estimation of stationary distributions because the approximation usually introduces a sink-state.

Truncations for stationary distributions often involve re-direction schemes for transitions leaving and entering the subset. A comprehensive survey of such state-space truncation methods can be found in [Kun+21b]. A popular method of identifying truncations is the construction of a suitable Lyapunov function. Beyond their use for establishing ergodicity [MT93; GBK14; Day+11], these functions can be used to obtain truncations, guaranteed to contain a certain amount of stationary probability mass [Day+11]. Using Lyapunov functions for the construction of truncations often leads to very conservative sets [MAK14]. Different approaches have been employed to find truncations: In Gupta, Mikelson, and Khammash [GMK17] SSA estimates are used to set up an increasing family of truncations.

MOMENT-BASED APPROXIMATION   Apart from approaches based on state-space truncations, moment-based approaches have been particularly popular recently [Ghu+17; DB18a; Kun+19a; SH17]. Such approaches are based on the fact that particular matrices of distributional moments such as mean and variance are positive semi-definite. Along with linear constraints stemming from the Kolmogorov equations [BBW16], a semi-definite program can be formulated and solved using existing tools. While this method is suited to compute bounds on both moments and subsets of the state-space, its application is limited, due to numerical issues inherent in the formulation [DB18a].

An approach where quantities are only described in terms of their magnitude has been proposed in Ceska and Kretínský [CK19]. This allows for an efficient qualitative analysis of both dynamic and transient behavior.

AGGREGATION   An aggregation scheme similar to the one used here has been previously proposed in Backenköhler et al. [Bac+21] to analyze the bridging problem on Markov population models. This is the problem of analyzing process dynamics under both initial and terminal constraints.

Aggregation-based numerical methods for computing the stationary distribution of discrete or continuous-time Markov chains

have been studied in previous work. Popular approaches rely on an alternation of aggregation and disaggregation of the state-space [Ste94; Sch91]. In the case of stiff chains, such aggregations are typically based on a separation of time-scales [CS85]. However, these methods have been developed for finite chains with arbitrary structure and are motivated by numerical issues of standard methods such as the power method or Jacobi iteration [Ste94]. They do not consider a truncation of irrelevant states, while here our aggregation approach is used to determine the most relevant states under stationary conditions in large or infinite chains with population structure.

6.2   TRUNCATION-BASED APPROXIMATION OF π_∞

In many relevant cases, the state-space is huge or infinite and therefore the stationary solution cannot be computed directly. To make such a computation possible we have to restrict ourselves to a finite manageable subset of the state-space and assume the majority of the probability mass is concentrated within that finite subset. The main problem is to deal with the transitions leading to and from the truncated set (cf. Figure 6.1). In forward analysis, the outgoing transitions are simply redirected into a sink-state. This way, a forward analysis provides lower bounds since mass leaving the truncation does not re-enter. This approach, however, is unsuitable for the computation of stationary distributions because mass would accumulate in the sink-state leading to a distribution assigning all mass to it. Therefore, transitions leaving the truncation need to be redirected back into the truncation.

The process' dynamics outside the truncation are defined by the *stochastic complement* [Spi14]. If its behavior was known, one could redirect outgoing to incoming transitions optimally and preserve the correct stationary distribution. However, this reentry distribution is typically unknown in most relevant cases. Many different reentry distributions have been used, such as redirecting to some internal state or states with incoming transition from outside the truncation. Reference [Kun+21a] provides a comprehensive review of such methods.

The most natural choice is to pick a reentry distribution that redirects mass to states with incoming transitions from truncated states (cf. Figure 6.1 (center)).

Using varying redirections, we can compute bounds on the stationary probability conditioned on a truncation [Spi14, (Thm. 14)]. To do this, one has to compute the stationary distribution for every possible way of connecting all outgoing to a single incoming transition. Naturally, such an algorithm is rather expensive since one has to solve a linear system for each combination. Therefore this method of computing bounds is costly on very large truncations, often given by Lyapunov functions.

When computing an approximation instead of bounds, we employ a uniform redirection scheme: Outgoing transitions are split uniformly among incoming transitions. Due to the threshold-based truncation scheme, we are likely to end up with a somewhat uniform distribution over in-boundary states (see Section 6.2.2).

The identification of good truncations remains a major task in such approximations. Using approaches such as Lyapunov functions (Section 2.6.1) [Day+11] or moment-bounds [Kun+21a] can provide a good initial estimate, but typically the resulting truncations are far larger than necessary. This leads to dramatically increased computational costs, especially when bounding methods mentioned above are performed. Until a system for a larger truncation is solved, the precise location of most of the probability mass is often unknown. Instead of solving the full system for such a large space, we employ an aggregation scheme to cover large areas of the state-space with exponentially fewer variables.

Error bounds have been derived for increasing truncation sets in the case of linear Lyapunov functions [GMK17]. However, until now it has not been shown that these bounds are applicable in practice [MT+94]. Alternatively, one can monitor the product of the probability-ouflow rate and the maximum L1-norm, which bounds the approximation error up to a constant $M > 0$, assuming a linear Lyapunov function exists [GMK17].
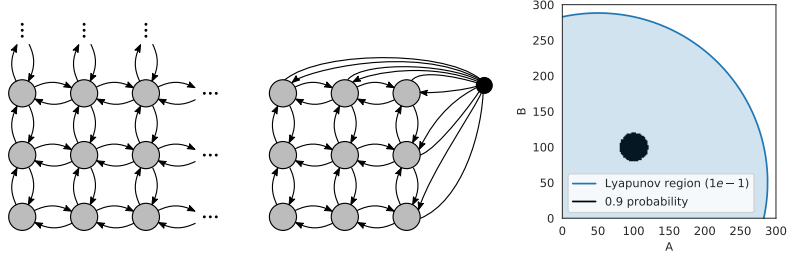
Figure 6.1: (left) A countably infinite state-space. (center) Outgoing transitions are re-directed (according to the reentry distribution) to states that have incoming transitions from outside the truncation. (right) A comparison of the area prescribed by a Lyapunov analysis using Geobound and threshold 0.1 and the minimal area containing 0.9 stationary probability mass. The model is a parallel birth-death process (Model 7).

6.2.1 *Initial Aggregation*

The initial aggregated space $\hat{\mathcal{S}}^{(0)}$ should encompass all regions of the state-space that could contain significant mass because states outside this initial area will not be refined. In principle, multiple approaches could be used to identify such a region. One possibility is the computation of moment bounds for the stationary distribution [Ghu+17; DB18a]. Based on these bounds on expectations and covariances, an initial truncation could be fixed. The approach we use here is to identify such a region by a Lyapunov analysis [Day+11]. This way, we obtain a polynomial describing a semi-algebraic subset of the entire state-space containing $1 - \epsilon_\ell$ of the mass, where $\epsilon_\ell > 0$ can be fixed arbitrarily. These sets usually are far larger than a minimal set containing $1 - \epsilon_\ell$ of stationary probability mass would be. As an initial aggregation, we build an aggregation on a subset $[0..n]^{n_S} \subset \mathcal{S}$ containing the set prescribed by the Lyapunov analysis. We also employ this approach to estimate errors in the evaluation. Specifically, we employ the tool Geobound [Spi10] with L2-norm as function g implementing techniques presented in Dayar et al. [Day+11].

In many cases, simple choices of g such as the L1- or L2- norm are sufficient. However, the sets resulting from such functions are often very conservative. Consider Figure 6.1 (right) as an example,
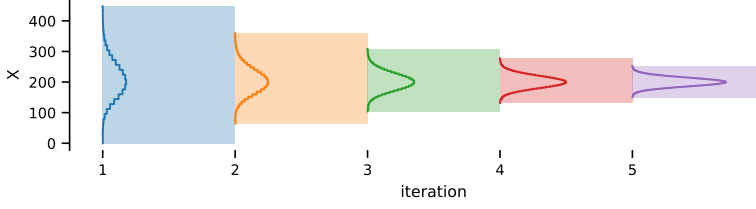
Figure 6.2: The state-space refinement algorithm on a birth-death process. From left to right the state size is halved and states with low probability are removed from the truncation. The final truncation is a typical truncation with states of size 1 and the initial states are of size 2^4.

where the Lyapunov truncation with $\epsilon_\ell = 0.1$ for two parallel birth-death processes (Model 7) is compared to the smallest set containing 0.9 of stationary probability. Clearly, the area given by the Lyapunov function is magnitudes larger than necessary to capture probability mass consistent with ϵ_ℓ.

6.2.2 *Iterative Refinement Algorithm*

Algorithm 3: Approximating the stationary distribution

input : Initial partitioning $S^{(0)}$, truncation threshold ϵ
output: approximate stationary distribution $\hat{\pi}_\infty$

1 for $i = 1, \ldots, m$ **do**

2 $\hat{\pi}_\infty^{(i)} \leftarrow$ solve approximate stationary distribution on $S^{(i)}$;

3 $\mathcal{R} \leftarrow$ smallest $\mathcal{R}' \subseteq S^{(i)}$ such that
$$\sum_{\bar{x} \in \mathcal{R}'} \hat{\pi}_\infty^{(i)}(\bar{x}) \geqslant 1 - \epsilon;$$

4 $S^{(i+1)} \leftarrow \bigcup_{\bar{x} \in \mathcal{R}} \text{split}(\bar{x})$;

5   update $\hat{Q}$-matrix;

6 return $\hat{\pi}_\infty^{(m)}$;

The refinement algorithm (Algorithm 3) starts with a set of large macro-states that are iteratively refined, based on approximate stationary distributions. We start by constructing square macro-states of size 2^m in each dimension for some $m \in \mathbb{N}$

such that they form a large-scale grid $S^{(0)}$. Hence, each initial macro-state has a volume of $(2^m)^{n_S}$. This choice of grid size is convenient because we can halve states in each dimension. Moreover, this choice ensures that all states have an equal volume and we end up with unit-sized macro-states, equivalent to a truncation of the original non-lumped state-space.

An iteration of the state-space refinement starts by computing the stationary distribution, using the lumped $\hat{Q}$-matrix. Based on a threshold parameter $\epsilon > 0$ states are either removed or split (line 4), depending on the mass assigned to them by the approximate stationary probabilities $\hat{\pi}_\infty^{(i)}$. Thus, each macro-state is either split into 2^{n_S} new states or removed entirely. The result forms the next lumped state-space $S^{(i+1)}$. The $\hat{Q}$-matrix is updated (line 5) using (5.4) to calculate the transition rates of the next aggregated truncation $S^{(i+1)}$. Entries of truncated states are removed from the updated transition matrix. Transitions leading to them are re-directed according to the re-entry matrix (Section 6.2). After m iterations (we started with states of side lengths 2^m) we have a standard FSP scheme on the original model tailored to computing an approximation of the stationary distribution.

This way, the refinement algorithm focuses only on those parts of the state-space contributing most to the stationary distribution. For instance, in Figure 6.2 the stationary probability mass mostly concentrates around $\#S = 200$. Therefore, states that are further away from this area can be dropped in further refinement. This filtering (line 3 in Algorithm 3) ensures that states contributing significantly to $\hat{\pi}_\infty^{(i)}$ will be kept and refined in the next iteration. The selection of states is done by sorting states in descending order according to their approximate probability mass. This ensures the construction of the smallest possible subset chosen for refinement according to the approximation. Then states are collected until their overall approximate mass is above $1 - \epsilon$.

An interesting feature of the aggregation scheme is that the distribution tends to spread out more. This is due to the assumption of a uniform distribution inside macro-states. To gain an intuition, consider a macro-state that encompasses a peak of the stationary distribution. If we re-distribute the actual probability mass inside this macro-state uniformly, a higher probability is assigned to states at the macro-state's border. When plugging

such macro-states together, this increased mass away from the peak will increase the mass assigned to adjacent macro-states. This effect is illustrated by the example of a birth-death process in Figure 6.2. Due to this effect, an iterative refinement typically keeps an over-approximation in terms of state-space area. This is a desirable feature since relevant regions are less likely to be pruned due to lumping approximations.

6.3 RESULTS

A prototype was implemented in Rust 1.50 and Python 3.8. The linear systems were solved either using Numpy [Har+20] for up to 5000 states, or the sparse linear solver as available through Scipy [Vir+20], or the iterative biconjugate gradient stabilized algorithm [Vor92] (up to 10,000 iterations and absolute tolerance 10^{-16}).

The examples that we consider in the sequel are typical benchmarks for the analysis of MPMs. For most of them, appropriate Lyapunov functions have been determined using Geobound [Spi10; Spi14]. However, the corresponding Lyapunov sets containing at least $1 - \epsilon_\ell$ of the stationary probability mass are very large for typical choices of ϵ_ℓ (e.g. $\epsilon_\ell \in \{0.1, 0.05, 0.001\}$). Even for extremely large ϵ_ℓ, say $\epsilon_\ell = 0.8$, the remaining state-space may still be huge (e.g, 15,198 states).

6.3.1 *Parallel Birth-Death Process*

We first examine the algorithm on the simple example of two parallel, uncoupled birth-death processes.

Model 7 (Parallel Birth-Death Process)**.** *Two uncoupled parallel birth-death processes result in a simple stationary distribution that is given by a product of two Poisson distributions.*

$$\varnothing \xrightarrow{\rho} A \qquad A \xrightarrow{\delta} \varnothing \qquad \varnothing \xrightarrow{\rho} B \qquad B \xrightarrow{\delta} \varnothing$$

As a parameterization we choose $\rho = 100$ and $\delta = 1$.

For this model, the stationary distribution is known to be the product of two Poisson distributions with rate ρ/δ.
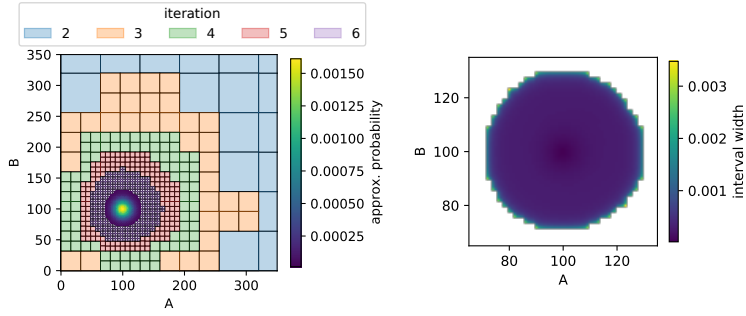
Figure 6.3: Results for Model 7 with truncation threshold $\epsilon = 0.1$. (left) Truncations of different iterations are layered on top of each other. At higher iterations, truncations cover less area but increase in detail, due to the refinement of macro-states. The final approximation is indicated by its approximate probabilities. (right) The difference between the upper and lower bounds on the probability conditioned on the truncation.

According to the Lyapunov analysis with a 1e-4 bound, we fix the initial truncation to a 70×70 grid of macro-states with size 2^7 in each dimension. This implies 8 iterations of the algorithm to arrive at a truncation with the original granularity. In Figure 6.3, we illustrate the truncations of different iterations. Over the iterations, the covered area decreases, while the aggregation granularity increases. The final truncation distribution approximation is also depicted and covers $1 - 1.27\text{e-}2$ of the true stationary distribution (cf. Table 6.2).

For this case study, we also compute state-wise bounds on the probabilities conditioned on the truncation as discussed in Section 6.2. In Figure 6.5 (right), we present the difference between upper and lower bound for $\epsilon = 0.1$. We observe intervals that are narrowest in the truncation's interior near the distribution's mode. The largest intervals or the largest absolute uncertainty is present in the boundary states. This indicates, that the specific reentry distribution has little effect on the main approximate stationary mass. More detailed results on the intervals' magnitudes are given in Table 6.2.

	threshold parameter ϵ			
	1e-1	1e-2	1e-3	1e-4
total width	1.2336	3.09e-02	5.39e-04	8.12e-06
max. width	3.47e-03	9.29e-05	4.04e-07	4.65e-09
outside mass	1.27e-02	1.05e-04	1.05e-06	1.06e-08

Table 6.1: Results for Model 7 : The characteristics of the lower-upper bound intervals on the conditional probability and the mass not contained in the truncation are given.
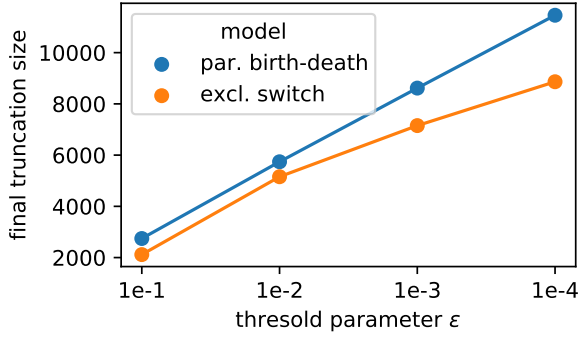

Figure 6.4: The sizes of the final truncation vs. the threshold parameter ϵ (Model 1 and Model 8).
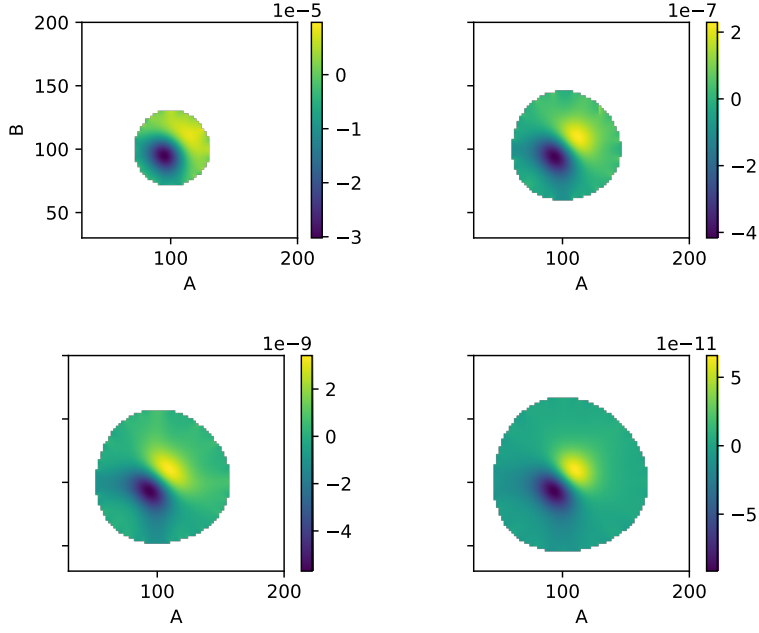
Figure 6.5: The error over the truncation wrt. the analytical solution

6.3.2 Exclusive Switch

The exclusive switch [BB08] has three different modes of operation, depending on the DNA state, i.e. on whether a protein of type one or two is bound to the DNA.

Model 8 (Exclusive Switch). *The exclusive switch model consists of a promoter region that can express both proteins P_1 and P_2. Both can bind to the region, suppressing the expression of the other protein. For certain parameterizations, this leads to a bi-modal or even tri-modal behavior.*

$$D \xrightarrow{\rho_1} D + P_1 \qquad D \xrightarrow{\rho_2} D + P_2 \qquad P_1 \xrightarrow{\lambda} \varnothing \qquad P_2 \xrightarrow{\lambda} \varnothing$$

$$D + P_1 \xrightarrow{\beta} D.P_1 \qquad D.P_1 \xrightarrow{\gamma_1} D + P_1 \qquad D.P_1 \xrightarrow{\rho_1} D.P_1 + P_1$$

$$D + P_2 \xrightarrow{\beta} D.P_2 \qquad D.P_2 \xrightarrow{\gamma_2} D + P_2 \qquad D.P_2 \xrightarrow{\rho_2} D.P_2 + P_2$$
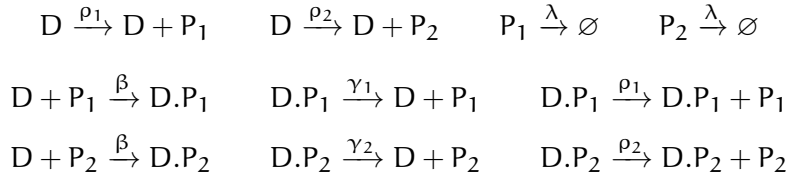
We choose parameter values $\rho_1 = 0.7$, $\rho_2 = 0.6$, $\lambda = 0.02$, $\beta = 0.005$, $\gamma_1 = 0.06$, and $\gamma_2 = 0.05$.

Since the exclusive switch models mutually exclusive binding of proteins to a single genetic locus, we know a priori that there are exactly three distinct operating modes. In particular are D, $D.P_1$, and $D.P_2$ mutually exclusive such that

$$X_D(t) + X_{D.P_1}(t) + X_{D.P_2}(t) = 1, \quad \forall t \geqslant 0.$$

This model characteristic often leads to bi-modal stationary distributions, where one or the other protein is more abundant depending on the genetic state.

Accordingly, we adjust the initial truncation: The state-space for the DNA states is not lumped. Instead we "stack" lumped approximations of the P_1-P_2 plane upon each other. Such special treatment of DNA states is common for such models [LMW11]. Using Lyapunov analysis for threshold 0.001, we fix an initial state-space of 63×63 macro-states with size 2^7. Detailed results for different parameters ϵ are presented Table 2. We compute error bounds using a worst-case analysis based on reference solutions provided by Geobound with $\epsilon_\ell = 0.01$. We observe a strong decrease in both upper bounds on the total absolute and maximal absolute error in the final iteration. Interestingly, the errors between different thresholds are very close in earlier iterations. This is mainly due to the usage of absolute errors which causes probabilities close to the mode dominate.

Using Geobound we observe that our final truncation captures the stationary mass very well (cf. Table 6.2). We use the Geobound's lower bounds with $\epsilon_\ell = 1e-2$ and find that the uncovered mass by the aggregation-based truncation is magnitudes lower than ϵ or close to it (for $\epsilon = 0.1$). While they capture the mass well, they are much smaller than the Geobound truncation ($\epsilon_\ell = 0.1$) with 16,780 states, regardless of the threshold parameter ϵ.

In Figure 6.4, we show the effect of the threshold parameter ϵ on the size of the final truncation. We observe a roughly linear increase in size with an exponential decrease of ϵ.

6.3.3 *p53 Oscillator*

We now consider a model of the interactions of the tumor suppressor p53 [GZ+06]. The system describes the negative feedback

	threshold parameter ϵ			
	1e-1	1e-2	1e-3	1e-4
total width	5.5171	1.5559	2.89e-02	3.71e-04
max. width	1.58e-01	3.30e-03	3.47e-05	3.84e-07
outside mass $\leqslant$	1.52e-01	1.29e-03	2.02e-05	2.72e-07

Table 6.2: Results for Model 8: The characteristics of the lower-upper bound intervals on the conditional probability and the upper bound on mass not contained in the truncation are given.
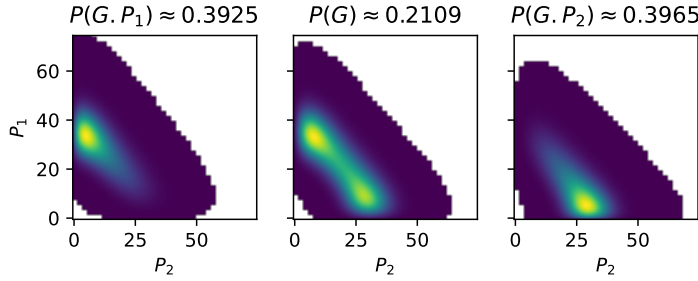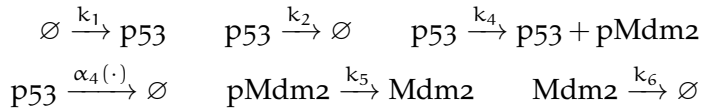

Figure 6.6: The approximate stationary distribution of the exclusive switch (Model 8) obtained with $\epsilon = $ 1e-4.

loop between p53 and the oncogene Mdm2. Species pMdm2 models a precursor to Mdm2. This model is particularly interesting due to its complex three-dimensional oscillatory behavior. The model is ergodic with a unique stationary distribution [GBK14].

Model 9 (p53 Oscillator).

$$\varnothing \xrightarrow{k_1} \text{p53} \qquad \text{p53} \xrightarrow{k_2} \varnothing \qquad \text{p53} \xrightarrow{k_4} \text{p53} + \text{pMdm2}$$

$$\text{p53} \xrightarrow{\alpha_4(\cdot)} \varnothing \qquad \text{pMdm2} \xrightarrow{k_5} \text{Mdm2} \qquad \text{Mdm2} \xrightarrow{k_6} \varnothing$$

The non-polynomial degradation reaction rate

$$\alpha_4(x) = k_3 x_{\text{Mdm2}} \frac{x_{\text{p53}}}{x_{\text{p53}} + k_7} .$$

The parameterization based on [AKS13] is $k_1 = 90$, $k_2 = 0.002$, $k_3 = 1.7$, $k_4 = 1.1$, $k_5 = 0.93$, $k_6 = 0.96$, *and* $k_7 = 0.01$.

With the exception of propensity function α_4, we can compute the transition rates $\bar{\alpha}_i$ using the Faulhaber formulae, as discussed in Chapter 5. We consider α_4 separately, because it is non-polynomial and therefore, we have to make an approximation. The fraction occurring in the non-linear propensity function α_4 can roughly be characterized as an activation function: Due to the low value of parameter $k_7 = 0.01$ we can approximate

$$\frac{x_{p53}}{x_{p53} + k_7} \approx \begin{cases} 0 & \text{if } x_{p53} = 0 \\ 1 & \text{otherwise} \end{cases}$$

Note, that $\sum_{i=0}^{n} i/(i+k_7)$ can be solved analytically. However, the approximation presented above is much simpler to compute.

We use this approximation at the coarser levels of aggregation to efficiently compute the approximate transition rate $\bar{\alpha}_4$. At the fines granularity we switch back to exact propensity function α_4.

We now derive Lyapunov-sets for the p53 oscillator case study (Model 9). Let the Lyapunov function

$$g(x) = 120x_{p53} + 0.2x_{pMdm2} + 0.1x_{Mdm2} . \tag{6.1}$$

Then the drift

$$\begin{aligned} d(x) = & -\frac{k_3 x_{Mdm2} x_{p53}}{x_{p53} + k_7} - 0.1k_6 x_{Mdm2} + 120k_1 \\ & - 120k_2 x_{p53} + 0.2k_4 x_{p53} - 0.1k_5 x_{pMdm2} \\ = & -\frac{204 x_{Mdm2} x_{p53}}{x_{p53} + 0.01} - 0.096 x_{Mdm2} - 0.02 x_{p53} \\ & - 0.0093 x_{pMdm2} + 10800 . \end{aligned} \tag{6.2}$$

Clearly, $c = \sup_{x \in S} d(x) = 10800$. In particular, the supremum c is at the origin since all non-constant terms are negative. The slowest rate of decrease for (6.2) is x_{p53} with $x_{Mdm2} = x_{pMdm2} = 0$. We are content with a superset of a Lyapunov set (2.27) for some threshold ϵ_ℓ. Therefore taking (2.27), we can solve the inequality

$$\frac{\epsilon_\ell}{c}(c - 0.02 x_{p53}) > \epsilon_\ell - 1$$

for x_{p53} and
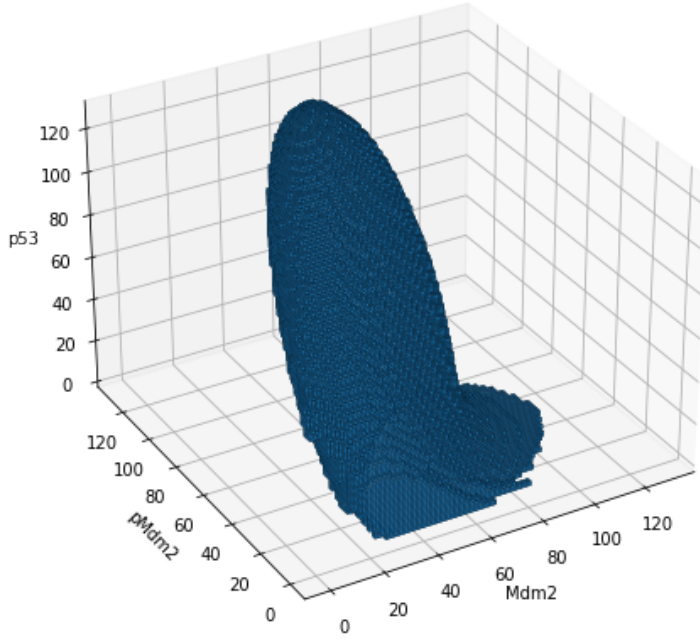
$$\frac{c}{0.02\epsilon_\ell} < x_{p53} . \tag{6.3}$$

Figure 6.7: The final truncation at original granularity derived for the p53 oscillator.

Therefore

$$\pi_\infty \left(\left\{ x \in \mathcal{S} \mid \frac{c}{0.2\epsilon_\ell} < \|x\| \right\} \right) > 1 - \epsilon_\ell . \qquad (6.4)$$

Due to the exponential increase stemming from the three-dimensional nature of this model, we only evaluated with parameter $\epsilon = 0.1$. According to the Lyapunov analysis shown above, the area covered by an $6 \times 6 \times 6$ macro-states with size 2^{20}, covers 0.9 of stationary mass. A truncation of this same area would consist of 226,492,416 states instead of the 216 macro-states. The model has a striking oscillatory behavior (cf. Figure 6.8) that is reflected in its stationary distribution. This feature is well-captured in the approximate distribution, where the oscillatory behavior leads to a complex stationary distribution (cf. Figure 6.9). This distribution leads to a non-trivial truncation (357,488 states) which is tailored to the main stationary mass (Figure 6.7).
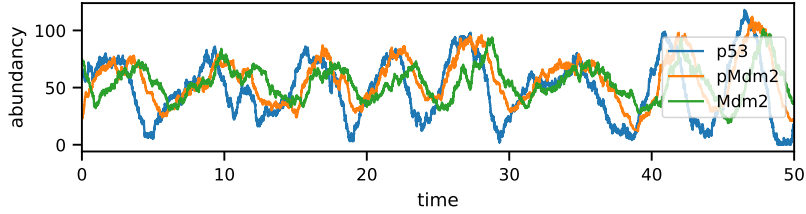
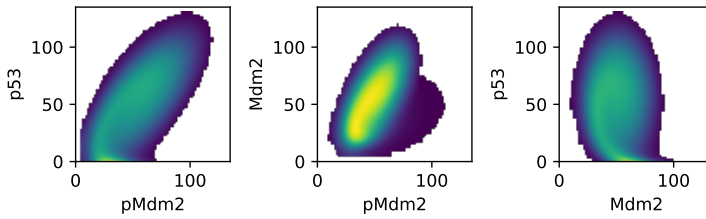Figure 6.8: A sample trajectory illustrating the oscillatory long-run behavior.

Figure 6.9: The approximate marginal distributions of the stationary distribution based on the truncation derived with $\epsilon = 0.1$.

6.4   CONCLUSION

State-of-the-art methods for numerically calculating the stationary distribution of Markovian population models rely on coarse truncations of irrelevant parts of large or infinite discrete state-spaces. These truncations are either obtained from the stationary statistical moments of the process or from Lyapunov theory. They are limited in shape because these methods do not take into account the detailed steady-state flow within the truncated state-space but only consider the average drift or stationary moments.

Here, we propose a method to find a tight truncation that is not limited in its shape and iteratively optimizes the set based on numerically cheap solutions of abstract intermediate models. It captures the main portion of probability mass even in the case of complex behaviors efficiently. In particular, the method represents another option, where Lyapunov analysis leads to forbiddingly large truncations.

ANALYSIS UNDER TERMINAL CONSTRAINTS

Many tasks, such as the analysis of rare events or the inference of agent counts under partial observations naturally introduce terminal constraints on the system. In these cases, the system's initial state is known, as well as the system's (partial) state at a later time-point. The probabilities corresponding to this so-called *bridging problem* are often referred to as *bridging probabilities* [GS19; GW11]. For instance, if the exact, full state of the process X_t has been observed at time $t = 0$ and $t = T$, the bridging distribution is given by

$$\Pr(X_t = x \mid X_0 = x_0, X_T = x_g)$$

for all states x and times $t \in [0, T]$. Often, the condition is more complex, such that in addition to an initial distribution, a terminal distribution is present. Such problems typically arise in a Bayesian setting, where the a priori behavior of a system is filtered such that the posterior behavior is compatible with noisy, partial observations [Bro17; Hua+16]. For example, time-series data of protein levels is available while the mRNA concentration is not [Ada+17; Hua+16]. In such a scenario our method can be used to identify a good truncation to analyze the probabilities of mRNA levels.

Bridging probabilities also appear in the context of rare events. Here, the rare event is the terminal constraint because we are only interested in paths containing the event. Typically researchers have to resort to Monte-carlo simulations in combination with variance reduction techniques in such cases [DJ+11; KM08].

Efficient numerical approaches that are not based on sampling or ad-hoc approximations have rarely been developed.

Here, we combine state-of-the-art truncation strategies based on a forward analysis [LMW11; And+11] with a refinement approach that starts from an abstract MPM with lumped states. We base this lumping on a grid-like partitioning of the state-space. Throughout a lumped state, we assume a uniform distribution that gives an efficient and convenient abstraction of

the original MPM. Note that the lumping does not follow the classical paradigm of Markov chain lumpability [Buc94] or its variants [DS97]. Instead of an approximate block structure of the transition-matrix used in that context, we base our partitioning on a segmentation of the molecule counts. Moreover, during the iterative refinement of our abstraction, we identify those regions of the state-space that contribute most to the bridging distribution. In particular, we refine those lumped states that have a bridging probability above a certain threshold δ and truncate all other macro-states. This way, the algorithm learns a truncation capturing most of the bridging probabilities. This truncation provides guaranteed lower bounds because it is at the granularity of the original model.

7.1 RELATED WORK

TERMINAL CONSTRAINTS   The problem of endpoint constrained analysis occurs in the context of Bayesian estimation [Sär13]. For MPMs, this problem has been addressed by Huang et al. [Hua+16] using moment closure approximations and by Wildner and Koeppl [WK19] further employing variational inference. Golightly and Sherlock modified stochastic simulation algorithms to approximatively augment generated trajectories [GS19]. Since a statistically exact augmentation is only possible for few simple cases, diffusion approximations [GW05] and moment approximations [MGW13] have been employed. Such approximations, however, do not give any guarantees on the approximation error and may suffer from numerical instabilities [SSG14].

RARE EVENT PROBABILITIES   The bridging problem also arises during the estimation of first passage times and rare event analysis. Approaches for first-passage times are often of heuristic nature [Sch+17; HSBl12; BL14]. Rigorous approaches yielding guaranteed bounds are currently limited by the performance of state-of-the-art optimization software [BBW20]. In biological applications, rare events of interest are typically related to the reachability of certain thresholds on molecule counts or mode switching [STM12]. Most methods for the estimation of rare event probabilities rely on importance sampling [KM08; DJ+11]. For

other queries, alternative variance reduction techniques such as control variates are available [BBW19]. Apart from sampling-based approaches, dynamic finite-state projections have been employed by Mikeev, Sandmann, and Wolf [MSW13], but are lacking automated truncation schemes.

STATE-SPACE TRUNCATION   The analysis of countably infinite state-spaces is often handled by a pre-defined truncation [KNP11]. Sophisticated state-space truncations for the (unconditioned) forward analysis have been developed to give lower bounds and rely on a trade-off between computational load and tightness of the bound [MK06; LMW11; And+11; HMW09; Mik+13].

REACHABILITY IN VERIFICATION   Reachability analysis, which is relevant in the context of probabilistic verification [BL14; Neu+19], is a bridging problem where the endpoint constraint is the visit of a set of goal states. Backward probabilities are commonly used to compute reachability likelihoods [AD13; ZK06]. Approximate techniques for reachability, based on moment closure and stochastic approximation, have also been developed in [BL14; BLN18], but lack error guarantees.

HIDDEN MARKOV MODELS   There is also a conceptual similarity between computing bridging probabilities and the forward-backward algorithm for computing state-wise posterior marginals in hidden Markov models (HMMs) [RJ86]. Like MPMs, HMMs are a generative model that can be conditioned on observations. We only consider two observations (initial and terminal state) that are not necessarily noisy but the forward and backward probabilities admit the same meaning.

7.2   BACKWARDS PROBABILITIES

Let $x_g \in \mathcal{S}$ be a fixed goal state. Given the terminal constraint

$$\Pr(X_T = x_g) = 1 \text{ for some } T \geqslant 0, \tag{7.1}$$

we are interested in the so-called backward probabilities

$$\beta(x_i, t) = \Pr(X_T = x_g \mid X_t = x_i), \quad t \leqslant T. \tag{7.2}$$

Note that $\beta(\cdot, t)$ is a function of the conditional event and thus is no probability distribution over the state-space. Instead $\beta(\cdot, t)$ gives the reaching probabilities for all states over the time span of $[t, T]$. To compute these probabilities, we can employ the Kolmogorov backward equation

$$\frac{d}{dt}\beta(t) = Q\beta(t), \tag{7.3}$$

where we use the same vectorization to construct $\beta(t)$ as we used for $\pi(t)$. The above equation is integrated backwards in time and yields the reachability probability for each state x_i and time $t < T$ of ending up in x_g at time T.

The state-space of many MPMs, even simple ones, is countably infinite. In this case, we have to truncate the state-space to a *reasonable* finite subset. The choice of this truncation heavily depends on the goal of the analysis. If one is interested in the most "common" behavior, for example, a dynamic mass-based truncation scheme is most appropriate [MS19]. Such a scheme truncates states with small probability during the numerical integration. However, common mass-based truncation schemes are not as useful for the bridging problem. This is because trajectories that meet the specific terminal constraints can be far off the main bulk of the probability mass. We solve this problem by a state-space lumping in connection with an iterative refinement scheme.

7.3   BRIDGING DISTRIBUTION

The process' probability distribution given both initial and terminal constraints is formally described by the conditional probabilities

$$\gamma(x_i, t) = \Pr(X_t = x_i \mid X_0 = x_0, X_T = x_g), \quad 0 \leqslant t \leqslant T \tag{7.4}$$

for fixed initial state x_0 and terminal state x_g. We call these probabilities the *bridging probabilities*. It is straight-forward to see that γ admits the factorization

$$\gamma(x_i, t) = \pi(x_i, t)\beta(x_i, t)/\pi(x_g, T) \tag{7.5}$$

due to the Markov property. The normalization factor, given by the reachability probability $\pi(x_g, T) = \beta(x_0, 0)$, ensures that

$\gamma(\cdot, t)$ is a distribution for all time points $t \in [0, T]$. We call each $\gamma(\cdot, t)$ a *bridging distribution*. From the Kolmogorov equations (2.17) and (7.3) we can obtain both the forward probabilities $\pi(\cdot, t)$ and the backward probabilities $\beta(\cdot, t)$ for $t < T$.

We can easily extend this procedure to deal with hitting times constrained by a finite time-horizon by making the goal state x_g absorbing.

EXAMPLE   In Figure 7.1 we plot the forward, backward, and bridging probabilities for Model 1. The probabilities are computed on a $[0, 100]$ state-space truncation. The approximate forward solution $\hat{\pi}$ shows how the probability mass drifts upwards towards the stationary distribution Poisson(100). The backward probabilities are highest for states below the goal state $x_g = 40$. This is expected because upwards drift makes reaching x_g more probable for "lower" states. Finally, the approximate bridging distribution $\hat{\gamma}$ can be recognized to be proportional to the product of forward $\hat{\pi}$ and backward probabilities $\hat{\beta}$.   ◇
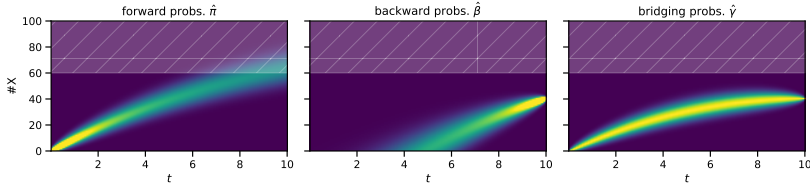
Figure 7.1: Forward, backward, and bridging probabilities for Model 1 with initial constraint $X_0 = 0$ and terminal constraint $X_{10} = 40$ on a truncated state-space. Probabilities over 0.1 in $\hat{\pi}$ and $\hat{\beta}$ are given full intensity for visual clarity. The lightly shaded area ($\geqslant 60$) indicates a region being more relevant for the forward than for the bridging probabilities.

7.4 BRIDGE TRUNCATION VIA LUMPING APPROXIMATIONS

We first discuss the truncation of countably infinite state-spaces to analyze backward and forward probabilities (Section 7.4.1). To identify effective truncations we employ a lumping scheme. Finally, in Section 7.4.2 we present an iterative refinement algorithm yielding a suitable truncation for the bridging problem.

7.4.1 *Finite State Projection*

Even in simple models such as a birth-death Process (Model 1), the reachable state-space is countably infinite. Direct analyzes of backward (7.2) and forward equations (2.16) are often infeasible. Instead, the integration of these differential equations requires working with a finite subset of the infinite state-space [MK06]. If states are truncated, their incoming transitions from states that are not truncated can be re-directed to a *sink state*. The accumulated probability in this sink state is then used as an error estimate for the forward integration scheme. Consequently, many truncation schemes, such as dynamic truncations [And+11], aim to minimize the amount of "lost mass" of the forward probability. We use the same truncation method but base the truncation on bridging probabilities rather than the forward probabilities.

7.4.2 *Iterative Refinement Algorithm*

The iterative refinement algorithm (Algorithm 4) starts with a set of large macro-states that are iteratively refined, based on approximate solutions to the bridging problem. We start by constructing square macro-states of size 2^m in each dimension for some $m \in \mathbb{N}$ such that they form a large-scale grid $S^{(0)}$. Hence, each initial macro-state has a volume of $(2^m)^{n_s}$. This choice of grid size is convenient because we can halve states in each dimension. Moreover, this choice ensures that all states have equal volume and we end up with states of volume $2^0 = 1$ which is equivalent to a truncation of the original non-lumped state-space.

An iteration of the state-space refinement starts by computing both the forward and backward probabilities (line 2 and line 3) via integration of (2.17) and (7.3), respectively, using the lumped $\hat{Q}$-matrix. Based on the resulting approximate forward and backward probabilities, we compute an approximation of the bridging distributions (line 4). This is done for each time-point in an equispaced grid on $[0, T]$. The time grid granularity is a hyper-parameter of the algorithm. If the grid is too fine, the memory overhead of storing backward $\hat{\beta}^{(i)}$ and forward solutions $\hat{\pi}^{(i)}$ increases. We denote the approximations with a hat (e.g. $\hat{\pi}$) rather than a bar (e.g. $\bar{\pi}$) to indicate that not only

the lumping approximation but also a truncation is applied and similarly for the Q-matrix. If, on the other hand, the granularity is too low, too much of the state-space might be truncated. Based on a threshold parameter $\delta > 0$ states are either removed or split (line 7), depending on the mass assigned to them by the approximate bridging probabilities $\hat{\gamma}_t^{(i)}$. A state can be split by the split-function which halves the state in each dimension. Otherwise, it is removed. Thus, each macro-state is either split into 2^{n_s} new states or removed entirely. The result forms the next lumped state-space $S^{(i+1)}$. The Q-matrix is adjusted (line 10) such that transition rates for $S^{(i+1)}$ are calculated according to (5.4). Entries of truncated states are removed from the transition matrix. Transitions leading to them are re-directed to a sink state (Section 7.4.1). After m iterations (we started with states of side lengths 2^m) we have a standard FSP scheme on the original model tailored to computing an approximation of the bridging distribution.

Algorithm 4: Iterative refinement for the bridging problem

    input : Initial partitioning $S^{(0)}$, truncation threshold δ
    output: approximate bridging distribution $\hat{\gamma}$

1 **for** $i = 1, \ldots, m$ **do**
2      $\hat{\pi}_t^{(i)} \leftarrow$ solve approximate forward equation on $S^{(i)}$;
3      $\hat{\beta}_t^{(i)} \leftarrow$ solve approximate backward equation on $S^{(i)}$;
4      $\hat{\gamma}_t^{(i)} \leftarrow \hat{\beta}^{(i)}\hat{\pi}^{(i)}/\hat{\pi}(x_g, T)$; /* approx. bridging dist. */
5      $S^{(i+1)} \leftarrow \emptyset$;
6   **foreach** $\bar{x} \in S^{(i)}$ **do**
7   **if** $\exists t. \hat{\gamma}_t^{(i)}(\bar{x}) \geqslant \delta$;   /* refinement */
8   **then**
9             $S^{(i+1)} \leftarrow S^{(i+1)} \cup \text{split}(\bar{x})$;
10   update $\hat{Q}$-matrix;
11 **return** $\hat{\gamma}^{(i)}$;

In Figure 7.2 we give a demonstration of how Algorithm 4 works to refine the state-space iteratively. Starting with an initial lumped state-space $S^{(0)}$ covering a large area of the state-space,
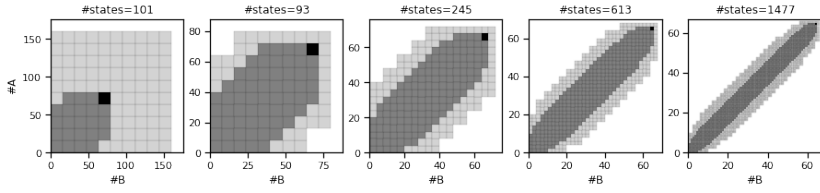
Figure 7.2: The state-space refinement algorithm on two parallel unit-rate arrival processes. The bridging problem from $(0, 0)$ to $(64, 64)$ and $T = 10$ and truncation threshold $\delta = 5\text{e-}3$. States with a bridging probability below δ are light grey. The macro-state containing the goal state is marked in black. The initial macro-states are of size 16×16.

repeated evaluations of the bridging distributions are performed. After five iterations the remaining truncation includes all states that significantly contribute to the bridging probabilities over the times $[0, T]$.

It is important to realize that determining the most relevant states is *the* main challenge. The above algorithm solves this problem by considering only those parts of the state-space that contribute most to the bridging probabilities. The truncation is tailored to this condition and might ignore regions that are likely in the unconditioned case. For instance, in Figure 7.1 the bridging probabilities mostly remain below a population threshold of $\#X = 60$ (as indicated by the lighter/darker coloring), while the forward probabilities mostly exceed this bound. Hence, in this example a significant portion of the forward probabilities $\hat{\pi}_t^{(i)}$ is captured by the sink state. However, the condition in line 7 of Algorithm 4 ensures that states contributing significantly to $\hat{\gamma}_t^{(i)}$ will be kept and refined in the next iteration.

7.5   RESULTS

We present four examples in this section to evaluate our proposed method. A prototype was implemented in Python 3.8. For numerical integration we used the Scipy implementation [Vir+20] of the implicit method based on backward-differentiation formulas [BH75]. The analysis as a Jupyter notebook is made available online [Bac20].

7.5.1 *Bounding Rare Event Probabilities*

We consider a simple model of two parallel Poisson processes describing the production of two types of agents. The corresponding probability distribution has Poisson product form at all time points $t \geqslant 0$ and hence we can compare the accuracy of our numerical results with the exact analytic solution. We use the proposed approach to compute lower bounds for rare event probabilities. [1]

Model 10 (Parallel Poisson Processes)**.** *The model consists of two parallel independent Poisson processes with unit rates.*

$$\varnothing \xrightarrow{1} A \qquad and \qquad \varnothing \xrightarrow{1} B \, .$$

The initial condition $X_0 = (0,0)$ *holds with probability one. After* t *time units each species abundance is Poisson distributed with rate* $\lambda = t$.

We consider the final constraint of reaching a state where both processes exceed a threshold of 64 at time 20. Without prior knowledge, a reasonable truncation would have been 160×160. But our analysis shows that just 20% of the states are necessary to capture over 99.6% of the probability mass reaching the target event (cf. Table 7.1). Decreasing the threshold δ leads to a larger set of states retained after truncation as more of the bridging distribution is included (cf. Figure 7.3). We observe an increase in truncation size that is approximately logarithmic in δ, which, in this example, indicates robustness of the method with respect to the choice of δ.

COMPARISON TO OTHER METHODS   The truncation approach that we apply is similar to the one used by Mikeev, Sandmann, and Wolf [MSW13] for rare event estimation. However, they used a given linearly biased MPM model to obtain a truncation. A general strategy to compute an appropriate biasing was not proposed. It is possible to adapt our truncation approach to the dynamic scheme in Mikeev, Sandmann, and Wolf [MSW13] where

---

[1] These bounds are rigorous up to the approximation error of the numerical integration scheme. However, the forward solution could be replaced by an adaptive uniformization approach [And+10] for a more rigorous integration error control.
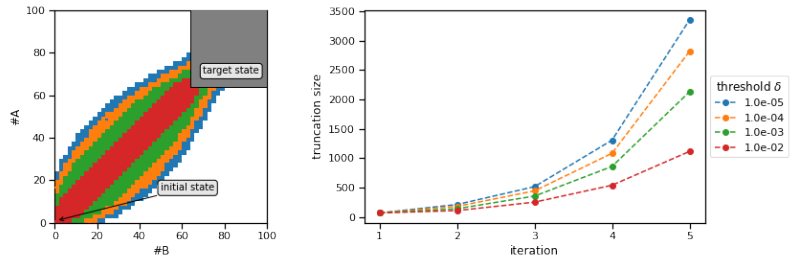
Figure 7.3: State-space truncation for varying values of the threshold parameter δ: Two parallel Poisson processes under terminal constraints $X_{20}^{(A)} \geqslant 64$ and $X_{20}^{(B)} \geqslant 64$. The initial macrostates are 16×16 such that the final states are regular micro states.

threshold δ	1e-2	1e-3	1e-4	1e-5
truncation size	1154	2354	3170	3898
overall states	2074	3546	4586	5450
estimate	8.88e-30	1.85e-29	1.86e-29	1.86e-29
rel. error	5.22e-01	3.66e-03	3.74e-05	9.52e-08

Table 7.1: Estimated reachability probabilities based on varying truncation thresholds δ: The true probability is 1.8625e-29. We also report the size of the final truncation and the accumulated size of all truncations during refinement iterations (overall states).

states are removed in an on-the-fly fashion during numerical integration.

A finite state-space truncation covering the same area as the initial lumping approximation would contain 25,600 states. The standard approach would be to build up the entire state-space for such a model [KNP11]. Even using a conservative truncation threshold δ = 1e-5, our method yields an accurate estimate using only about a fifth (5450) of this accumulated over all intermediate lumped approximations.

The goal is not treated as a single state. Otherwise, it consisted of 24,130 states.
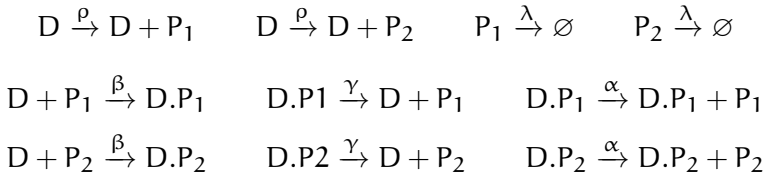
7.5.2   *Mode Switching*

Mode switching occurs in models exhibiting *multi-modal* behavior (see Section 2.7 for details) when a trajectory traverses a potential barrier from one mode to another. Often, mode switching is a rare event and occurs in the context of gene regulatory networks where a mode is characterized by the set of genes being currently active [Loi+07]. Similar dynamics also commonly occur in queuing models where a system may for example switch its operating behavior stochastically if traffic increases above or decreases below certain thresholds. Using the presented method, we can get both a qualitative and quantitative understanding of switching behavior without resorting to Monte-Carlo methods such as importance sampling.

This often is an instance of rare-event analysis.

7.5.2.1   *Exclusive Switch*

The exclusive switch [BB08] has three different modes of operation, depending on the deoxyribonucleic acid (DNA) state, i.e. on whether a protein of type one or two is bound to the DNA.

Model 11 (Exclusive Switch). *The exclusive switch model consists of a promoter region that can express both proteins P_1 and P_2. Both can bind to the region, suppressing the expression of the other protein. For certain parameterizations, this leads to a bi-modal or even tri-modal behavior.*

$$D \xrightarrow{\rho} D + P_1 \qquad D \xrightarrow{\rho} D + P_2 \qquad P_1 \xrightarrow{\lambda} \varnothing \qquad P_2 \xrightarrow{\lambda} \varnothing$$

$$D + P_1 \xrightarrow{\beta} D.P_1 \qquad D.P1 \xrightarrow{\gamma} D + P_1 \qquad D.P_1 \xrightarrow{\alpha} D.P_1 + P_1$$

$$D + P_2 \xrightarrow{\beta} D.P_2 \qquad D.P2 \xrightarrow{\gamma} D + P_2 \qquad D.P_2 \xrightarrow{\alpha} D.P_2 + P_2$$

The parameter values are $\rho = 1e\text{-}1$, $\lambda = 1e\text{-}3$, $\beta = 1e\text{-}2$, $\gamma = 8e\text{-}3$, and $\alpha = 1e\text{-}1$.

Since we know a priori of the three distinct operating modes, we adjust the method slightly: The state-space for the DNA states is not lumped. Instead we "stack" lumped approximations of the P_1-P_2 phase space upon each other. Special treatment of DNA states is common for such models [LMW11].

To analyze the switching, we choose the transition from

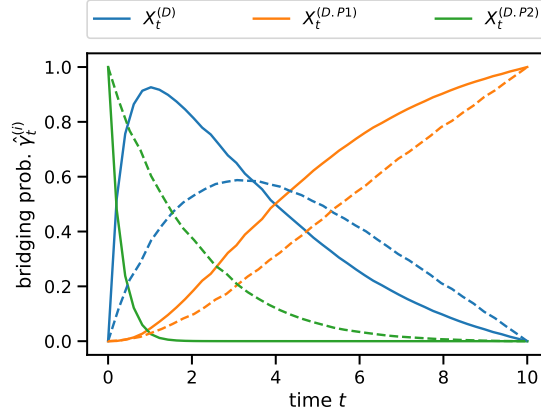*Variable order: P_1, P_2, D, D.P_1, D.P_2.*

Figure 7.4: Mode probabilities of the exclusive switch bridging problem over time for the first lumped approximation (dashed lines) and the final approximation (solid lines) with constraints $X_0 = (32, 0, 0, 1, 0)$ and $X_{10} = (0, 32, 0, 0, 1)$.

$$x_1 = (32, 0, 0, 0, 1) \quad \text{to} \quad x_2 = (0, 32, 0, 1, 0)$$

over the time interval $t \in [0, 10]$. The initial lumping scheme covers up to 80 molecules of P_1 and P_2 for each mode. Macrostates have size 8×8 and the truncation threshold is $\delta = 1e\text{-}4$.

In the analysis of biological switches, not only the switching probability but also the switching dynamics is a central part of understanding the underlying biological mechanisms. In Figure 7.4, we therefore plot the time-varying probabilities of the gene state conditioned on the mode. We observe a rapid unbinding of P_2, followed by a slow increase of the binding probability for P_1. These dynamics are already qualitatively captured by the first lumped approximation (dashed lines).

7.5.2.2 *Toggle Switch*

Next, we apply our method to a toggle switch model exhibiting non-polynomial rate functions. This well-known model considers two proteins A and B inhibiting the production of the respective other protein [Lip+06].
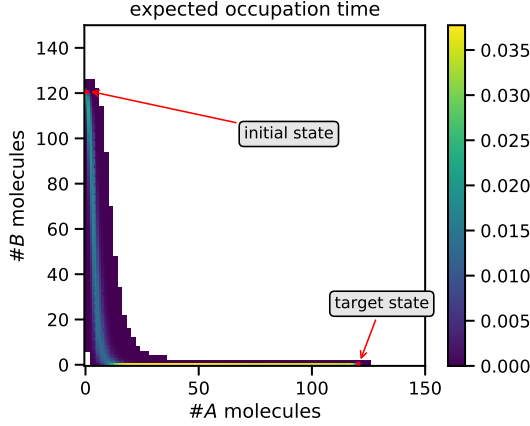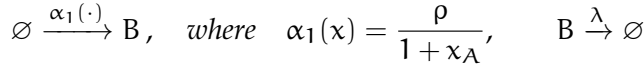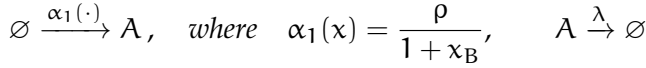
Figure 7.5: The expected occupation time (excluding initial and terminal states) for the switching problem of the toggle switch using Hill-type functions. The bridging problem is from initial $(0, 120)$ to a first passage of $(120, 0)$ in $t \in [0, 10]$.

Model 12 (Toggle Switch using Hill functions). *We have population types* A *and* B *with the following reactions and reaction rates.*

$$\varnothing \xrightarrow{\alpha_1(\cdot)} A, \quad where \quad \alpha_1(x) = \frac{\rho}{1 + x_B}, \quad A \xrightarrow{\lambda} \varnothing$$

$$\varnothing \xrightarrow{\alpha_1(\cdot)} B, \quad where \quad \alpha_1(x) = \frac{\rho}{1 + x_A}, \quad B \xrightarrow{\lambda} \varnothing$$

The parameterization is $\rho = 10$, $\lambda = 0.1$.

Due to the non-polynomial rate functions α_1 and α_2, the transition rates between macro-states are approximated by using the continuous integral

$$\bar{\alpha}_1(\bar{x}) \approx \int_{a-0.5}^{b+0.5} \frac{\rho}{1 + x} \, dx = \rho \left(\log\left(b + 1.5\right) - \log\left(a + 0.5\right) \right)$$

for a macro-state $\bar{x} = \{a, \dots, b\}$.

We analyze the switching scenario from $(0, 120)$ to the first visit of state $(120, 0)$ up to time $T = 10$. The initial lumping scheme covers up to 352 molecules of A and B and macro-states have size 32×32. The truncation threshold is $\delta = 1\text{e-}4$. The resulting truncation is shown in Figure 7.5. It also illustrates

the kind of insights that can be obtained from the bridging distributions. For an overview of the switching dynamics, we look at the expected occupation time under the terminal constraint of having entered state $(120, 0)$. Letting the corresponding hitting time be $\tau = \inf\{t \geqslant 0 \mid X_t = (120, 0)\}$, the expected occupation time for some state x is

$$E\left(\int_0^\tau 1_{=x}(X_t)\,dt \mid \tau \leqslant 10\right).$$

We observe that in this example the switching behavior seems to be asymmetrical. The main mass seems to pass through an area where initially a small number of A molecules is produced followed by a total decay of B molecules.

7.5.3   *Recursive Bayesian Estimation*

We now turn to the method's application in recursive Bayesian estimation. This is the problem of estimating the system's past, present, and future behavior under given observations. Thus, the MPM becomes a hidden Markov model (HMM). The observations in such models are usually noisy, meaning that we cannot infer the system state with certainty.

This estimation problem entails more general distributional constraints on terminal $\beta(\cdot, T)$ and initial $\pi(\cdot, 0)$ distributions than the point mass distributions considered up until now. We can easily extend the forward and backward probabilities to more general initial distributions and terminal distributions $\beta(T)$. For the forward probabilities we get

$$\pi(x_i, t) = \sum_j \Pr(X_t = x_i \mid X_0 = x_j)\pi(x_j, 0), \tag{7.6}$$

and similarly the backward probabilities are given by

$$\beta(x_i, t) = \sum_j \Pr(X_T = x_j \mid X_t = x_i)\beta_T(x_j). \tag{7.7}$$

We apply our method to an susceptible-exposed-infected-removed (SEIR) model. This is widely used to describe the spreading of an epidemic such as the current COVID-19 outbreak [HPS20; GBW20]. Temporal snapshots of the epidemic spread are mostly

only available for a subset of the population and suffer from inaccuracies of diagnostic tests. Bayesian estimation can then be used to infer the spreading dynamics given uncertain temporal snapshots.

Model 13 (Epidemics Model). *A population of susceptible individuals can contract a disease from infected agents. In this case, they are exposed, meaning they will become infected but cannot yet infect others. After being infected, individuals change to the removed state. The mass-action reactions are as follows.*

$$S + I \xrightarrow{\lambda} E + I \qquad E \xrightarrow{\mu} I \qquad I \xrightarrow{\rho} R$$
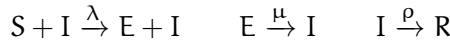
The parameter values are $\lambda = 0.5$, $\mu = 3$, $\rho = 3$. Due to the stoichiometric invariant $X_t^{(S)} + X_t^{(E)} + X_t^{(I)} + X_t^{(R)} = \text{const.}$, we can eliminate R from the system.

We consider the following scenario: We know that initially $(t = 0)$ one individual is infected and the rest is susceptible. At time $t = 0.3$ all individuals are tested for the disease. The test, however, only identifies infected individuals with probability 0.99. Moreover, the probability of a false positive is 0.05. We like to identify the distribution given both the initial state and the measurement at time $t = 0.3$. In particular, we want to infer the distribution over the latent counts of S and E by *recursive Bayesian estimation*.

The posterior for n_I infected individuals at time t, given measurement $Y_t = \hat{n}_I$ can be computed using Bayes' rule

$$\Pr(X_t^{(I)} = n_I \mid Y_t = \hat{n}_I) \propto \Pr(Y_t = \hat{n}_I \mid X_t^{(I)} = n_I) \Pr(X_t^{(I)} = n_I).$$
(7.8)

This problem is an extension of the bridging problem discussed up until now. The difference is that the terminal posterior is estimated it using the result of the lumped forward equation and the measurement distribution using (7.8). Based on this estimated terminal posterior, we compute the bridging probabilities and refine the truncation tailored to the location of the posterior distribution. In Figure 7.6a, we illustrate the bridging distribution between the terminal posterior and initial distribution. In the

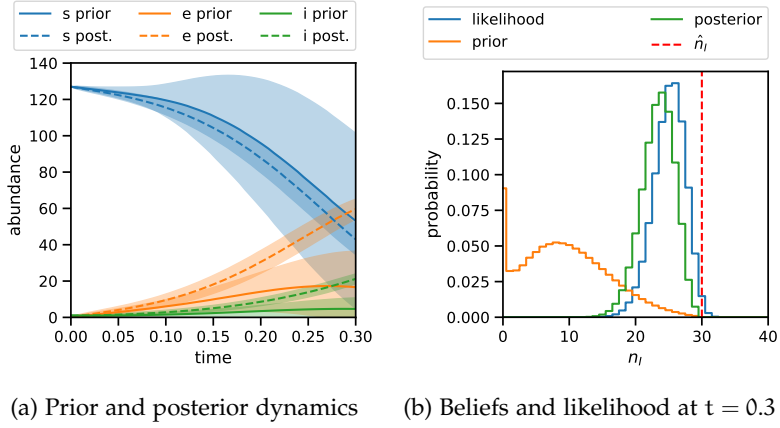(a) Prior and posterior dynamics   (b) Beliefs and likelihood at t = 0.3

Figure 7.6: (a) A comparison of the prior dynamics and the posterior smoothing (bridging) dynamics. (b) The prior, likelihood, and posterior of the number of infected individuals n_I at time t = 0.3 given the measurement $\hat{n}_I = 30$.
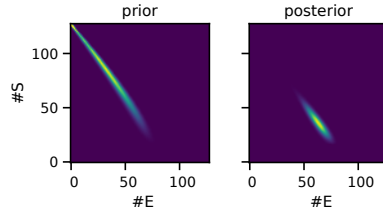
Figure 7.7: The prior and posterior distribution over the latent types E and S at time t = 0.3.

context of filtering problems this is commonly referred to as smoothing. Using the learned truncation, we can obtain the posterior distribution for the number of infected individuals at t = 0.3 (Figure 7.6b). Moreover, can we infer a distribution over the unknown number of susceptible and exposed individuals (Figure 7.7).

7.6   CONCLUSION

The analysis of Markov Jump processes with constraints on the initial and terminal behavior is an important part of many

probabilistic inference tasks such as parameter estimation using Bayesian or maximum likelihood estimation, inference of latent system behavior, the estimation of rare event probabilities, and reachability analysis for the verification of temporal properties. If endpoint constraints correspond to atypical system behaviors, standard analysis methods fail as they have no strategy to identify those parts of the state-space relevant for meeting the terminal constraint.

Here, we proposed a method that is not based on stochastic sampling and statistical estimation but provides a direct numerical approach. It starts with an abstract lumped model, which is iteratively refined such that only those parts of the model are considered that contribute to the probabilities of interest. In the final step of the iteration, we operate at the granularity of the original model and compute lower bounds for these bridging probabilities that are rigorous up to the error of the numerical integration scheme.

Our method exploits the population structure of the model, which is present in many important application fields of MPMs. Based on experience with other work based on truncation, the approach can be expected to scale up to at least a few million states [MSW11]. Compared to previous work, our method neither relies on approximations of unknown accuracy nor additional information such as a suitable change of measure in the case of importance sampling. It only requires a truncation threshold and an initial choice for the macro-state sizes.

In future work, we plan to extend our method to hybrid approaches, in which a moment representation is employed for large populations while discrete counts are maintained for small populations. Moreover, we will apply our method to model checking where constraints are described by some temporal logic [Haj+19].

Part IV

CONCLUDING REMARKS

8

CONCLUSION

Part V

APPENDIX

ADDITIONAL RESULTS & DATA

	ε		1	2	3	4	5	6	7	8
		$|S^{(i)}|$	4,900	28	52	112	232	472	960	1,932
1e-1		tot. error	1.91	1.84	1.73	1.55	1.29	9.35e-1	4.88e-1	3.54e-2
		max. error	3.15e-3	3.13e-3	3.08e-3	2.98e-3	2.77e-3	2.38e-3	1.57e-3	6.04e-5
		$|S^{(i)}|$	4,900	52	104	208	464	988	2,008	4,052
1e-2		tot. error	1.91	1.84	1.73	1.56	1.30	9.46e-1	5.01e-1	6.22e-4
		max. error	3.15e-3	3.13e-3	3.08e-3	2.98e-3	2.78e-3	2.39e-3	1.59e-3	8.33e-7
		$|S^{(i)}|$	4,900	84	152	300	652	1,440	2,996	6,068
1e-3		tot. error	1.91	1.83	1.73	1.56	1.30	9.46e-1	5.01e-1	9.83e-6
		max. error	3.15e-3	3.13e-3	3.08e-3	2.98e-3	2.78e-3	2.39e-3	1.59e-3	1.14e-8
		$|S^{(i)}|$	4,900	116	212	400	848	1,872	3,960	8,060
1e-4		tot. error	1.91	1.83	1.73	1.56	1.30	9.46e-1	5.01e-1	9.83e-6
		max. error	3.15e-3	3.13e-3	3.08e-3	2.98e-3	2.78e-3	2.39e-3	1.59e-3	1.83e-10

Table 1: Detailed results for Model 7. The errors are computed wrt. the reference Poissonian product. The total absolute error and the maximum absolute errors are given.

ϵ		iteration i							
		1	2	3	4	5	6	7	8
1e-1	$|S^{(i)}|$	11907	20	32	60	140	340	840	2116
	tot. error $\leqslant$	1.86e0	1.85e0	1.45e0	1.18e0	9.31e-1	6.41e-1	4.67e-1	4.89e-1
	max. error $\leqslant$	1.63e-3	1.63e-3	1.55e-3	1.40e-3	1.22e-3	9.36e-4	8.40e-4	1.40e-3
1e-2	$|S^{(i)}|$	11907	48	112	148	300	720	1892	5156
	tot. error $\leqslant$	1.86e0	1.84e0	1.44e0	1.21e0	9.56e-1	6.65e-1	3.41e-1	3.31e-2
	max. error $\leqslant$	1.63e-3	1.62e-3	1.53e-3	1.39e-3	1.20e-3	9.59e-4	5.86e-4	5.37e-5
1e-3	$|S^{(i)}|$	11907	84	192	244	488	1084	2692	7152
	tot. error $\leqslant$	1.86e0	1.83e0	1.46e0	1.22e0	9.63e-1	6.67e-1	3.37e-1	8.01e-4
	max. error $\leqslant$	1.63e-3	2.95e-2	1.54e-3	1.39e-3	1.20e-3	9.51e-4	5.79e-4	1.09e-6
1e-4	$|S^{(i)}|$	11907	124	324	352	672	1436	3408	8864
	tot. error $\leqslant$	1.86e0	1.83e0	1.46e0	1.22e0	9.63e-1	6.67e-1	3.37e-1	1.12e-5
	max. error $\leqslant$	1.63e-3	3.19e-2	1.54e-3	1.39e-3	1.20e-3	9.51e-4	5.79e-4	1.28e-8

Table 2: Detailed results for Model 8. Upper bounds on the total absolute error and the maximum absolute error are given. The worst-case errors are computed wrt. the reference Geobound solution with $\epsilon_\ell = 1e-2$.

Case study	n_λ	ϕ	$1 - \frac{\sigma_1^2}{\sigma_0^2}$	slowdown	efficiency	$|P|$
Dimerization	10	ϕ_{sq}	0.881641	1.530137	5.558692	1.621917
		ϕ_c	0.965224	1.945588	14.859417	3.338501
		ϕ_q	0.916445	1.625232	7.409904	1.997045
		ϕ_ℓ	0.868288	1.380344	5.529745	1.081152
	20	ϕ_{sq}	0.941153	1.637272	10.437978	1.842971
		ϕ_c	0.964204	1.907999	14.747328	2.915082
		ϕ_q	0.947984	1.747519	11.072422	2.227250
		ϕ_ℓ	0.931030	1.433401	10.169570	1.088572
	30	ϕ_{sq}	0.959517	1.723449	14.404936	1.972426
		ϕ_c	0.962514	1.770936	15.142156	2.216103
		ϕ_q	0.966216	1.847441	16.117387	2.446661
		ϕ_ℓ	0.945724	1.456432	12.710196	1.084188
Dist. mod.	10	ϕ_{sq}	0.619560	1.488483	1.770218	3.232575
		ϕ_c	0.700255	2.241695	1.492171	8.008607
		ϕ_q	0.643550	1.613001	1.743500	3.817641
		ϕ_ℓ	0.596650	1.459405	1.703170	2.657000
	20	ϕ_{sq}	0.697414	1.519425	2.181687	2.631677
		ϕ_c	0.713445	2.706546	1.292838	10.295856
		ϕ_q	0.697654	1.585313	2.092817	3.398235
		ϕ_ℓ	0.695846	1.473976	2.235418	2.226530
	30	ϕ_{sq}	0.712941	1.543068	2.263644	2.378037
		ϕ_c	0.721354	2.874249	1.252541	10.910880
		ϕ_q	0.711877	1.607712	2.164485	2.979704
		ϕ_ℓ	0.669963	1.522184	1.996300	2.085473
Excl. switch	10	ϕ_{sq}	0.807184	1.227471	4.239255	2.536479
		ϕ_c	0.880285	1.633530	5.135205	7.411732
		ϕ_q	0.849082	1.312416	5.067770	3.639250
		ϕ_ℓ	0.783459	1.195821	3.874778	2.090101
	20	ϕ_{sq}	0.856593	1.263340	5.539683	2.206154
		ϕ_c	0.910480	1.864405	6.011256	9.441336
		ϕ_q	0.867987	1.317958	5.765884	3.140806
		ϕ_ℓ	0.825518	1.243075	4.627662	1.981143
	30	ϕ_{sq}	0.869165	1.298893	5.905196	2.059415
		ϕ_c	0.921019	1.966191	6.461331	9.928998
		ϕ_q	0.876822	1.340409	6.079876	2.762449
		ϕ_ℓ	0.843288	1.288925	4.968796	1.983174

Table 3: $n_{max} = 1$, $n = 10{,}000$, $d = 100$, $k_{min} = 3$

Case study	n_λ	ϕ	$1 - \frac{\sigma_1^2}{\sigma_0^2}$	slowdown	efficiency	$|P|$
Dimerization	10	ϕ_{sq}	0.905254	1.659799	6.402491	2.388472
		ϕ_c	0.987526	2.474939	33.074955	6.501180
		ϕ_q	0.923063	1.822654	7.195544	3.179257
		ϕ_ℓ	0.878232	1.415909	5.830248	1.092264
	20	ϕ_{sq}	0.949038	1.831995	10.797164	2.890898
		ϕ_c	0.985710	2.391457	29.704344	5.450299
		ϕ_q	0.968076	2.021487	15.662368	3.681229
		ϕ_ℓ	0.925413	1.449386	9.298961	1.072761
	30	ϕ_{sq}	0.964855	1.924268	14.911787	3.026275
		ϕ_c	0.981507	2.144089	25.520987	4.179125
		ϕ_q	0.973902	2.095985	18.507746	3.685851
		ϕ_ℓ	0.948349	1.507425	12.904707	1.074538
Dist. mod.	10	ϕ_{sq}	0.619450	1.734737	1.519168	3.148184
		ϕ_c	0.665361	3.301159	0.909443	13.456259
		ϕ_q	0.680592	1.840457	1.705876	3.864240
		ϕ_ℓ	0.612674	1.662962	1.556868	2.659592
	20	ϕ_{sq}	0.684789	1.811408	1.755652	2.687379
		ϕ_c	0.689835	4.455005	0.726640	17.609554
		ϕ_q	0.687665	1.901258	1.688449	3.413595
		ϕ_ℓ	0.651262	1.770238	1.623924	2.266729
	30	ϕ_{sq}	0.690602	1.922217	1.686011	2.375455
		ϕ_c	0.649191	4.837419	0.591701	19.145054
		ϕ_q	0.701253	2.001179	1.677062	3.007525
		ϕ_ℓ	0.639123	1.894074	1.467403	2.086275
Excl. switch	10	ϕ_{sq}	0.811956	1.505521	3.544783	2.323999
		ϕ_c	0.916866	4.507566	2.681363	21.692390
		ϕ_q	0.868874	1.776190	4.309354	4.739893
		ϕ_ℓ	0.795802	1.466579	3.353046	2.016196
	20	ϕ_{sq}	0.832562	1.657484	3.617313	2.085711
		ϕ_c	0.934280	6.348223	2.406431	29.976320
		ϕ_q	0.878944	1.879341	4.416281	3.990881
		ϕ_ℓ	0.837922	1.647329	3.759896	1.978017
	30	ϕ_{sq}	0.829427	1.844766	3.190308	2.043201
		ϕ_c	0.947324	7.130628	2.673225	32.513670
		ϕ_q	0.878830	2.053317	4.034987	3.611746
		ϕ_ℓ	0.824936	1.838879	3.118728	1.978836

Table 4: $n_{max} = 2$, $n = 10,000$, $d = 100$, $k_{min} = 3$

BIBLIOGRAPHY

[Ada+17]   Aysun Adan, Günel Alizada, Yağmur Kiraz, Yusuf Baran, and Ayten Nalbant. "Flow cytometry: basic principles and applications." In: *Critical reviews in biotechnology* 37.2 (2017), pp. 163–176.

[AKS13]   Angelique Ale, Paul Kirk, and Michael PH Stumpf. "A general moment expansion method for stochastic kinetic models." In: *The Journal of chemical physics* 138.17 (2013), p. 174101.

[AD13]   Elvio Gilberto Amparore and Susanna Donatelli. "Backward Solution of Markov Chains and Markov Regenerative Processes: Formalization and Applications." In: *Electron. Notes Theor. Comput. Sci.* 296 (2013), pp. 7–26.

[ACK10]   David F Anderson, Gheorghe Craciun, and Thomas G Kurtz. "Product-form stationary distributions for deficiency zero chemical reaction networks." In: *Bulletin of mathematical biology* 72.8 (2010), pp. 1947–1970.

[AK11]   David F Anderson and Thomas G Kurtz. "Continuous time Markov chain models for chemical reaction networks." In: *Design and analysis of biomolecular circuits*. Springer, 2011, pp. 3–42.

[AY18]   David F Anderson and Chaojie Yuan. "Low variance couplings for stochastic models of intracellular processes with time-dependent rate functions." In: *Bulletin of mathematical biology* (2018), pp. 1–29.

[And12]   William J Anderson. *Continuous-time Markov chains: An applications-oriented approach*. Springer Science & Business Media, 2012.

[And+10]   Aleksandr Andreychenko, Pepijn Crouzen, Linar Mikeev, and Verena Wolf. "On-the-fly uniformization of time-inhomogeneous infinite Markov population models." In: *arXiv preprint arXiv:1006.4425* (2010).

[And+11]   Aleksandr Andreychenko, Linar Mikeev, David Spieler, and Verena Wolf. "Parameter identification for Markov models of biochemical reactions." In: *International Conference on Computer Aided Verification*. Springer. 2011, pp. 83–98.

[Azi+96]   Adnan Aziz, Kumud Sanwal, Vigyan Singhal, and Robert Brayton. "Verifying continuous time Markov chains." In: *International Conference on Computer Aided Verification*. Springer. 1996, pp. 269–276.

[Bac19]   Michael Backenköhler. *Stochastic Simulation Code*. https://github.com/mbackenkoehler/cme-simulation. 2019.

[Bac20]   Michael Backenköhler. *MJP Aggregation and Bridging Code*. 2020. URL: https://www.github.com/mbackenkoehler/mjp_bridging.

[Bac+21]   Michael Backenköhler, Luca Bortolussi, Gerrit Großmann, and Verena Wolf. "Analysis of Markov Jump Processes under Terminal Constraints." In: *27th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*. Vol. 1265. Lecture Notes in Computer Science. Springer, 2021, pp. 210–229.

[BBW16]   Michael Backenköhler, Luca Bortolussi, and Verena Wolf. "Generalized method of moments for stochastic reaction networks in equilibrium." In: *International Conference on Computational Methods in Systems Biology*. Springer. 2016, pp. 15–29.

[BBW18]   Michael Backenköhler, Luca Bortolussi, and Verena Wolf. "Moment-Based Parameter Estimation for Stochastic Reaction Networks in Equilibrium." In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 15.4 (2018), pp. 1180–1192.

[BBW19]   Michael Backenköhler, Luca Bortolussi, and Verena Wolf. "Control Variates for Stochastic Simulation of Chemical Reaction Networks." In: *17th International Conference on Computational Methods in Systems Biology (CMSB)*. Vol. 11773. Lecture Notes in Computer Science. Springer, 2019, pp. 42–59.

[BBW20]   Michael Backenköhler, Luca Bortolussi, and Verena Wolf. "Bounding Mean First Passage Times in Population Continuous-Time Markov Chains." In: *17th International Conference on Quantitative Evaluation of SysTems (QEST)*. Vol. 12289. Lecture Notes in Computer Science. Springer, 2020, pp. 155–174.

[Bai+03]   Christel Baier, Boudewijn Haverkort, Holger Hermanns, and J-P Katoen. "Model-checking algorithms for continuous-time Markov chains." In: *IEEE Transactions on software engineering* 29.6 (2003), pp. 524–541.

[Bai+00]   Christel Baier, Boudewijn Haverkort, Holger Hermanns, and Joost-Pieter Katoen. "Model checking continuous-time Markov chains by transient analysis." In: *International Conference on Computer Aided Verification*. Springer. 2000, pp. 358–372.

[BB08]   Baruch Barzel and Ofer Biham. "Calculation of switching times in the genetic toggle switch and other bistable systems." In: *Physical Review E* 78.4 (2008), p. 041919.

[BMN09]   Golan Bel, Brian Munsky, and Ilya Nemenman. "The simplicity of completion time dis-

tributions for common complex biochemical processes." In: *Physical biology* 7.1 (2009), p. 016003.

[BDNH16]   Marco Bernardo, Rocco De Nicola, and Jane Hillston, eds. *Formal Methods for the Quantitative Evaluation of Collective Adaptive Systems*. Vol. 9700. Lecture Notes in Computer Science. Springer, 2016.

[Bog+15]   Sergiy Bogomolov, Thomas A Henzinger, Andreas Podelski, Jakob Ruess, and Christian Schilling. "Adaptive moment closure for parameter inference of biochemical reaction networks." In: *International Conference on Computational Methods in Systems Biology*. Springer. 2015, pp. 77–89.

[Bor+13a]   Luca Bortolussi, Jane Hillston, Diego Latella, and Mieke Massink. "Continuous approximation of collective system behaviour: A tutorial." In: *Performance Evaluation* 70.5 (May 2013), pp. 317–349. (Visited on 06/11/2015).

[Bor+13b]   Luca Bortolussi, Jane Hillston, Diego Latella, and Mieke Massink. "Continuous approximation of collective system behaviour: A tutorial." In: *Performance Evaluation* 70.5 (2013), pp. 317–349.

[BHL20]   Luca Bortolussi, Jane Hillston, and Michele Loreti. "Fluid approximation of broadcasting systems." In: *Theoretical Computer Science* 816 (2020), pp. 221–248.

[BL13]   Luca Bortolussi and Roberta Lanciani. "Model checking Markov population models by central limit approximation." In: *International Conference on Quantitative Evaluation of Systems*. Springer. 2013, pp. 123–138.

[BL14]   Luca Bortolussi and Roberta Lanciani. "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. 2014, pp. 224–239.

[BLN18]   Luca Bortolussi, Roberta Lanciani, and Laura Nenzi. "Model checking Markov population models by stochastic approximations." In: *Inf. Comput.* 262 (2018), pp. 189–220.

[BMS15]   Luca Bortolussi, Dimitrios Milios, and Guido Sanguinetti. "Efficient stochastic simulation of systems with multiple time scales via statistical abstraction." In: *International Conference on Computational Methods in Systems Biology*. Springer. 2015, pp. 40–51.

[Bre03]   Lothar Breuer. *From Markov jump processes to spatial queues*. Springer, 2003.

[Bro17]   Lyle D Broemeling. *Bayesian Inference for Stochastic Processes*. CRC Press, 2017.

[Buc94]   Peter Buchholz. "Exact and ordinary lumpability in finite Markov chains." In: *Journal of applied probability* (1994), pp. 59–75.

[BH75]   George D. Byrne and Alan C. Hindmarsh. "A polyalgorithm for the numerical solution of ordinary differential equations." In: *ACM Transactions on Mathematical Software (TOMS)* 1.1 (1975), pp. 71–96.

[CS85]   Wei-Lu Cao and William J Stewart. "Iterative aggregation/disaggregation techniques for nearly uncoupled Markov chains." In: *Journal of the ACM (JACM)* 32.3 (1985), pp. 702–719.

[CGP05]   Yang Cao, Daniel T Gillespie, and Linda R Petzold. "The slow-scale stochastic simulation algorithm." In: *The Journal of chemical physics* 122.1 (2005), p. 014116.

[CCN12]   Luca Cardelli and Attila Csikász-Nagy. "The cell cycle switch computes approximate majority." In: *Scientific reports* 2 (2012), p. 656.

[CK19]   Milan Ceska and Jan Kretínský. "Semi-quanti-
            tative Abstraction and Analysis of Chemical
            Reaction Networks." In: *Computer Aided Verifi-
            cation - 31st International Conference, CAV 2019,
            New York City, NY, USA, July 15-18, 2019, Pro-
            ceedings, Part I.* 2019, pp. 475–496.

[Che+11]   Taolue Chen, Marco Diciolla, Marta Kwiatkow-
            ska, and Alexandru Mereacre. "Time-bounded
            verification of CTMCs against real-time spec-
            ifications." In: *International Conference on For-
            mal Modeling and Analysis of Timed Systems.*
            Springer. 2011, pp. 26–42.

[Che+09]   Taolue Chen, Tingting Han, Joost-Pieter Ka-
            toen, and Alexandru Mereacre. "Quantita-
            tive model checking of continuous-time Mar-
            kov chains against timed automata specifi-
            cations." In: *2009 24th Annual IEEE Sympo-
            sium on Logic In Computer Science.* IEEE. 2009,
            pp. 309–318.

[Che78]   Russell CH Cheng. "Analysis of simulation
            experiments under normality assumptions."
            In: *Journal of the Operational Research Society*
            29.5 (1978), pp. 493–497.

[DJ+11]   Bernie J Daigle Jr, Min K Roh, Dan T Gillespie,
            and Linda R Petzold. "Automated estimation
            of rare event probabilities in biochemical sys-
            tems." In: *The Journal of Chemical Physics* 134.4
            (2011), 01B628.

[Dav+15]   Alexandre David, Kim G Larsen, Axel Legay,
            Marius Mikučionis, Danny Bøgsted Poulsen,
            and Sean Sedwards. "Statistical model check-
            ing for biological systems." In: *International
            Journal on Software Tools for Technology Transfer*
            17.3 (2015), pp. 351–367.

[Day+11]   Tuğrul Dayar, Holger Hermanns, David Spie-
            ler, and Verena Wolf. "Bounding the equilib-
            rium distribution of Markov population mod-

els." In: *Numerical linear algebra with applications* 18.6 (2011), pp. 931–946.

[DS97]   Tuğrul Dayar and William J Stewart. "Quasi lumpability, lower-bounding coupling matrices, and nearly completely decomposable Markov chains." In: *SIAM Journal on Matrix Analysis and Applications* 18.2 (1997), pp. 482–498.

[Deh+17]   Christian Dehnert, Sebastian Junges, Joost-Pieter Katoen, and Matthias Volk. "A storm is coming: A modern probabilistic model checker." In: *International Conference on Computer Aided Verification*. Springer. 2017, pp. 592–600.

[DB16]   Steven Diamond and Stephen Boyd. "CVXPY: A Python-Embedded Modeling Language for Convex Optimization." In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[DB18a]   Garrett R Dowdy and Paul I Barton. "Bounds on stochastic chemical kinetic systems at steady state." In: *The Journal of chemical physics* 148.8 (2018), p. 084106.

[DB18b]   Garrett R Dowdy and Paul I Barton. "Dynamic bounds on stochastic chemical kinetic systems using semidefinite programming." In: *The Journal of chemical physics* 149.7 (2018), p. 074103.

[Eng06]   Stefan Engblom. "Computing the moments of high dimensional solutions of the master equation." In: *Applied Mathematics and Computation* 180.2 (2006), pp. 498–515.

[EK09]   Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. Vol. 282. John Wiley & Sons, 2009.

[GBT19]   Nicolas Gast, Luca Bortolussi, and Mirco Tribastone. "Size expansions of mean field approximation: Transient and steady-state analysis." In: *Performance Evaluation* 129 (2019), pp. 60–80.

[GZ+06]   Naama Geva-Zatorsky, Nitzan Rosenfeld, Sha-
           lev Itzkovitz, Ron Milo, Alex Sigal, Erez Dekel,
           Talia Yarnitzky, Yuvalal Liron, Paz Polak, Galit
           Lahav, et al. "Oscillations and variability in
           the p53 system." In: *Molecular systems biology*
           2.1 (2006), pp. 2006–0033.

[GLS18]   Khem Raj Ghusinga, Andrew Lamperski, and
           Abhyudai Singh. "Estimating stationary char-
           acteristic functions of stochastic systems via
           semidefinite programming." In: *2018 European
           Control Conference (ECC)*. IEEE. 2018, pp. 2720–
           2725.

[Ghu+17]   Khem Raj Ghusinga, Cesar A Vargas-Garcia,
           Andrew Lamperski, and Abhyudai Singh. "Ex-
           act lower and upper bounds on stationary mo-
           ments in stochastic biochemical systems." In:
           Physical biology 14.4 (2017), 04LT01.

[GS]   II Gihman and A Skorohod. *The Theory of Sto-
           chastic Processes II. 1975*.

[Gil77]   Daniel T Gillespie. "Exact stochastic simula-
           tion of coupled chemical reactions." In: *The
           journal of physical chemistry* 81.25 (1977), pp. 2340–
           2361.

[Gil00]   Daniel T Gillespie. "The chemical Langevin
           equation." In: *The Journal of Chemical Physics*
           113.1 (2000), pp. 297–306.

[Gil01]   Daniel T Gillespie. "Approximate accelerated
           stochastic simulation of chemically reacting
           systems." In: *The Journal of Chemical Physics*
           115.4 (2001), pp. 1716–1733.

[GY05]   Paul Glasserman and Bin Yu. "Large sam-
           ple properties of weighted monte carlo esti-
           mators." In: *Operations Research* 53.2 (2005),
           pp. 298–312.

[GS19]   Andrew Golightly and Chris Sherlock. "Effi-
            cient sampling of conditioned Markov jump
            processes." In: *Statistics and Computing* 29.5
            (2019), pp. 1149–1163.

[GW05]   Andrew Golightly and Darren J Wilkinson.
            "Bayesian inference for stochastic kinetic mod-
            els using a diffusion approximation." In: *Bio-
            metrics* 61.3 (2005), pp. 781–788.

[GW11]   Andrew Golightly and Darren J Wilkinson.
            "Bayesian parameter inference for stochastic
            biochemical network models using particle
            Markov chain Monte Carlo." In: *Interface focus*
            1.6 (2011), pp. 807–820.

[GBW20]   Gerrit Großmann, Michael Backenköhler, and
            Verena Wolf. "Importance of Interaction Struc-
            ture and Stochasticity for Epidemic Spread-
            ing: A COVID-19 Case Study." In: *17th Inter-
            national Conference on Quantitative Evaluation
            of SysTems (QEST)*. Vol. 12289. Lecture Notes
            in Computer Science. Springer, 2020, pp. 211–
            229.

[GBK14]   A Gupta, C Briat, and M Khammash. "A scal-
            able computational framework for establish-
            ing long-term behavior of stochastic reaction
            networks." In: *PLoS Comput Biol* 10.6 (2014),
            e1003669.

[GMK17]   Ankit Gupta, Jan Mikelson, and Mustafa Kham-
            mash. "A finite state projection algorithm for
            the stationary solution of the chemical master
            equation." In: *The Journal of chemical physics*
            147.15 (2017), p. 154101.

[Haj+19]   Matej Hajnal, Morgane Nouvian, David Šafrá-
            nek, and Tatjana Petrov. "Data-Informed Pa-
            rameter Synthesis for Population Markov Chains."
            In: *International Workshop on Hybrid Systems
            Biology*. Springer. 2019, pp. 147–164.

[Har+20]   Charles R. Harris et al. "Array programming with NumPy." In: *Nature* 585 (2020), pp. 357–362.

[Has+14]   Jan Hasenauer, Verena Wolf, Atefeh Kazeroonian, and Fabian J Theis. "Method of conditional moments (MCM) for the chemical master equation." In: *Journal of mathematical biology* 69.3 (2014), pp. 687–735.

[HSBl12]   Richard A Hayden, Anton Stefanek, and Jeremy T Bradley. "Fluid computation of passage-time distributions in large Markov models." In: *Theoretical Computer Science* 413.1 (2012), pp. 106–141.

[HPS20]   Shaobo He, Yuexi Peng, and Kehui Sun. "SEIR modeling of the COVID-19 and its dynamics." In: *Nonlinear Dynamics* (2020), pp. 1–14.

[HRS01]   Kurt Helmes, Stefan Röhl, and Richard H Stockbridge. "Computing moments of the exit time distribution for Markov processes by linear programming." In: *Operations Research* 49.4 (2001), pp. 516–530.

[HMW09]   Thomas A Henzinger, Maria Mateescu, and Verena Wolf. "Sliding window abstraction for infinite Markov chains." In: *International Conference on Computer Aided Verification*. Springer. 2009, pp. 337–352.

[Hes08]   Joao Hespanha. "Moment closure for biochemical networks." In: *2008 3rd International Symposium on Communications, Control and Signal Processing*. IEEE. 2008, pp. 142–147.

[Hin+06]   Andrew Hinton, Marta Kwiatkowska, Gethin Norman, and David Parker. "PRISM: A tool for automatic verification of probabilistic systems." In: *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer. 2006, pp. 441–444.

[Hua+16]   Lirong Huang, Loic Pauleve, Christoph Zechner, Michael Unger, Anders Hansen, and Heinz Koeppl. "Reconstructing dynamic molecular states from single-cell time series." In: *Journal of The Royal Society Interface* 13.122 (2016), p. 20160533.

[IBZ16]   Srividya Iyer-Biswas and Anton Zilman. "First-Passage Processes in Cellular Biology." In: *Advances in Chemical Physics* 160 (2016), pp. 261–306.

[JH07]   Tobias Jahnke and Wilhelm Huisinga. "Solving the chemical master equation for monomolecular reaction systems analytically." In: *Journal of mathematical biology* 54.1 (2007), pp. 1–26.

[KK09]   Kenji Kashima and Reiichiro Kawai. "Polynomial programming approach to weak approximation of Lévy-driven stochastic differential equations with application to option pricing." In: *2009 ICCAS-SICE*. IEEE. 2009, pp. 3902–3907.

[KTH14]   Atefeh Kazeroonian, Fabian J Theis, and Jan Hasenauer. "Modeling of stochastic biological processes with non-polynomial propensities using non-central conditional moment equation." In: *IFAC Proceedings Volumes* 47.3 (2014), pp. 1729–1735.

[Knu93]   Donald E Knuth. "Johann Faulhaber and sums of powers." In: *Mathematics of Computation* 61.203 (1993), pp. 277–294.

[Kun+19a]   Juan Kuntz, Philipp Thomas, Guy-Bart Stan, and Mauricio Barahona. "Bounding the stationary distributions of the chemical master equation via mathematical programming." In: *The Journal of chemical physics* 151.3 (2019), p. 034109.

[Kun+19b]   Juan Kuntz, Philipp Thomas, Guy-Bart Stan, and Mauricio Barahona. "The exit time finite state projection scheme: bounding exit distributions and occupation measures of continuous-time Markov chains." In: *SIAM Journal on Scientific Computing* 41.2 (2019), A748–A769.

[Kun+21a]   Juan Kuntz, Philipp Thomas, Guy-Bart Stan, and Mauricio Barahona. "Approximations of Countably Infinite Linear Programs over Bounded Measure Spaces." In: *SIAM Journal on Optimization* 31.1 (2021), pp. 604–625.

[Kun+21b]   Juan Kuntz, Philipp Thomas, Guy-Bart Stan, and Mauricio Barahona. "Stationary distributions of continuous-time Markov chains: a review of theory and truncation-based approximations." In: *SIAM Review* 63.1 (2021), pp. 3–64.

[Kur+18]   Pavel Kurasov, Alexander Lück, Delio Mugnolo, and Verena Wolf. "Stochastic hybrid models of gene regulatory networks–a PDE approach." In: *Mathematical biosciences* 305 (2018), pp. 170–177.

[Kur81]   Thomas G Kurtz. *Approximation of population processes*. Vol. 36. Regional conference series in applied mathematics. SIAM, 1981.

[KM08]   Hiroyuki Kuwahara and Ivan Mura. "An efficient and exact stochastic simulation method to analyze rare events in biochemical systems." In: *The Journal of chemical physics* 129.16 (2008), 10B619.

[KNP11]   Marta Kwiatkowska, Gethin Norman, and David Parker. "PRISM 4.0: Verification of probabilistic real-time systems." In: *International conference on computer aided verification*. Springer. 2011, pp. 585–591.

[L'E94]   Pierre L'Ecuyer. "Efficiency improvement and variance reduction." In: *Proceedings of the 26th conference on Winter simulation*. Society for Computer Simulation International. 1994, pp. 122–132.

[LMW11]   Maksim Lapin, Linar Mikeev, and Verena Wolf. "SHAVE: stochastic hybrid analysis of Markov population models." In: *Proceedings of the 14th international conference on Hybrid systems: computation and control*. 2011, pp. 311–312.

[Las10]   Jean-Bernard Lasserre. *Moments, positive polynomials and their applications*. Vol. 1. World Scientific, 2010.

[LPRZ06]   Jean-Bernard Lasserre, Tomas Prieto-Rumeau, and Mihail Zervos. "Pricing a class of exotic options via moments and SDP relaxations." In: *Mathematical Finance* 16.3 (2006), pp. 469–494.

[LMW82]   Stephen S Lavenberg, Thomas L Moeller, and Peter D Welch. "Statistical results on control variables with application to queueing network simulation." In: *Operations Research* 30.1 (1982), pp. 182–202.

[Lip+06]   Azi Lipshtat, Adiel Loinger, Nathalie Q Balaban, and Ofer Biham. "Genetic toggle switch without cooperative binding." In: *Physical review letters* 96.18 (2006), p. 188101.

[Loi+07]   Adiel Loinger, Azi Lipshtat, Nathalie Q Balaban, and Ofer Biham. "Stochastic simulations of genetic switch systems." In: *Physical Review E* 75.2 (2007), p. 021904.

[MOSEK ApS18] MOSEK ApS. *MOSEK Optimizer API for C 8.1.0.67*. 2018. URL: https://docs.mosek.com/8.1/capi/index.html.

[Mat+10]   M Mateescu, V Wolf, F Didier, and TA Henzinger. "Fast adaptive uniformisation of the chemical master equation." In: *IET systems biology* 4.6 (2010), pp. 441–452.

[MHK14]   Bence Mélykúti, Joao P Hespanha, and Mustafa Khammash. "Equilibrium distributions of simple biochemical reaction systems for time-scale separation in stochastic reaction networks." In: *Journal of The Royal Society Interface* 11.97 (2014), p. 20140054.

[MT93]   Sean P Meyn and Richard L Tweedie. "Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes." In: *Advances in Applied Probability* (1993), pp. 518–548.

[MT12]   Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer, 2012.

[MT+94]   Sean P Meyn, Robert L Tweedie, et al. "Computable bounds for geometric convergence rates of Markov chains." In: *The Annals of Applied Probability* 4.4 (1994), pp. 981–1011.

[Mik+13]   Linar Mikeev, Martin R Neuhäußer, David Spieler, and Verena Wolf. "On-the-fly verification and optimization of DTA-properties for large Markov chains." In: *Formal Methods in System Design* 43.2 (2013), pp. 313–337.

[MS19]   Linar Mikeev and Werner Sandmann. "Approximate Numerical Integration of the Chemical Master Equation for Stochastic Reaction Networks." In: *arXiv preprint arXiv:1907.10245* (2019).

[MSW11]   Linar Mikeev, Werner Sandmann, and Verena Wolf. "Efficient calculation of rare event probabilities in Markovian queueing networks." In: *Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools*. 2011, pp. 186–196.

[MSW13]   Linar Mikeev, Werner Sandmann, and Ver-
            ena Wolf. "Numerical approximation of rare
            event probabilities in biochemically reacting
            systems." In: *International Conference on Com-
            putational Methods in Systems Biology*. Springer.
            2013, pp. 5–18.

[MAK14]   Andreas Milias-Argeitis and Mustafa Kham-
            mash. "Optimization-based Lyapunov func-
            tion construction for continuous-time Markov
            chains with affine transition rates." In: *53rd
            IEEE Conference on Decision and Control*. IEEE.
            2014, pp. 4617–4622.

[MGW13]   Peter Milner, Colin S Gillespie, and Darren J
            Wilkinson. "Moment closure based parame-
            ter inference of stochastic kinetic models." In:
            Statistics and Computing 23.2 (2013), pp. 287–
            295.

[MS00]   Charles J Mode and Candace K Sleeman. *Sto-
            chastic processes in epidemiology: HIV/AIDS, other
            infectious diseases, and computers*. World Scien-
            tific, 2000.

[MK06]   Brian Munsky and Mustafa Khammash. "The
            finite state projection algorithm for the so-
            lution of the chemical master equation." In:
            The Journal of chemical physics 124.4 (2006),
            p. 044104.

[MNB09]   Brian Munsky, Ilya Nemenman, and Golan
            Bel. "Specificity and completion time distribu-
            tions of biochemical processes." In: *The Journal
            of chemical physics* 131.23 (2009), 12B616.

[Nel90]   Barry L Nelson. "Control variate remedies."
            In: *Operations Research* 38.6 (1990), pp. 974–
            992.

[Neu+19]   Thakur Neupane, Chris J Myers, Curtis Mad-
            sen, Hao Zheng, and Zhen Zhang. "STAMINA:
            STochastic approximate model-checker for IN-
            finite-state analysis." In: *International Confer-*

ence on Computer Aided Verification. Springer. 2019, pp. 540–549.

[O'D+17]   B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd. *SCS: Splitting Conic Solver, version 2.1.0.* https://github.com/cvxgrp/scs. Nov. 2017.

[Øks05]   Bernt Øksendal. *Stochastic differential equations: an introduction with applications*. 6th ed. Springer, 2005.

[Par08]   Etienne Pardoux. *Markov processes and applications: algorithms, networks, genome and finance*. Vol. 796. John Wiley & Sons, 2008.

[Par03]   Pablo A Parrilo. "Semidefinite programming relaxations for semialgebraic problems." In: *Mathematical programming* 96.2 (2003), pp. 293–320.

[PG16]   Mason A Porter and James P Gleeson. "Dynamical systems on networks." In: *Frontiers in Applied Dynamical Systems: Reviews and Tutorials* 4 (2016).

[RJ86]   Lawrence Rabiner and B Juang. "An introduction to hidden Markov models." In: *IEEE ASSP Magazine* 3.1 (1986), pp. 4–16.

[SH17]   Yuta Sakurai and Yutaka Hori. "A convex approach to steady state moment analysis for stochastic chemical reactions." In: *Decision and Control (CDC), 2017 IEEE 56th Annual Conference on*. IEEE. 2017, pp. 1206–1211.

[SH19]   Yuta Sakurai and Yutaka Hori. "Bounding Transient Moments of Stochastic Chemical Reactions." In: *IEEE Control Systems Letters* 3.2 (2019), pp. 290–295.

[Sär13]   Simo Särkkä. *Bayesian filtering and smoothing*. Vol. 3. Cambridge University Press, 2013.

[Sch+17]   David Schnoerr, Botond Cseke, Ramon Grima, and Guido Sanguinetti. "Efficient Low-Order Approximation of First-Passage Time Distributions." In: *Phys. Rev. Lett.* 119 (21 2017), p. 210601.

[SSG14]   David Schnoerr, Guido Sanguinetti, and Ramon Grima. "Validity conditions for moment closure approximations in stochastic chemical kinetics." In: *The Journal of chemical physics* 141.8 (2014), 08B616_1.

[SSG15]   David Schnoerr, Guido Sanguinetti, and Ramon Grima. "Comparison of different moment-closure approximations for stochastic chemical kinetics." In: *The Journal of Chemical Physics* 143.18 (Nov. 2015), p. 185101.

[SSG17]   David Schnoerr, Guido Sanguinetti, and Ramon Grima. "Approximation and inference methods for stochastic biochemical kinetics—a tutorial review." In: *Journal of Physics A: Mathematical and Theoretical* 50.9 (Mar. 2017), p. 093001.

[Sch91]   Paul J Schweitzer. "A survey of aggregation-disaggregation in large Markov chains." In: *Numerical solution of Markov chains* 8 (1991), pp. 63–88.

[SG+11]   Dan Siegal-Gaskins, Maria Katherine Mejia-Guerra, Gregory D Smith, and Erich Grotewold. "Emergence of switch-like behavior in a large family of simple biochemical networks." In: *PLoS Comput Biol* 7.5 (2011), e1002039.

[SH06]   Abhyudai Singh and Joao Pedro Hespanha. "Lognormal moment closures for biochemical reactions." In: *Proceedings of the 45th IEEE Conference on Decision and Control*. IEEE. 2006, pp. 2063–2068.

[Spi14]   D Spieler. "Numerical analysis of long-run properties for Markov population models." PhD thesis. Saarland University, 2014.

[Spi10]   David Spieler. *Geobound*. 2010. URL: https://mosi.uni-saarland.de/tools/geobound.

[SHZ11]   David Spieler, Ernst Moritz Hahn, and Lijun Zhang. "Model checking CSL for Markov population models." In: *arXiv preprint arXiv:1111.4385* (2011).

[SJ08]   Dov J Stekel and Dafyd J Jenkins. "Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression." In: *BMC systems biology* 2.1 (2008), p. 6.

[Ste94]   William J Stewart. *Introduction to the numerical solution of Markov chains*. Princeton University Press, 1994.

[Ste09]   William J Stewart. *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton university press, 2009.

[STM12]   Michael Strasser, Fabian J Theis, and Carsten Marr. "Stability and multiattractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression." In: *Biophysical journal* 102.1 (2012), pp. 19–29.

[Sze03]   Roberto Szechtman. "Control variate techniques for monte carlo simulation: control variates techniques for monte carlo simulation." In: *Proceedings of the 35th conference on Winter simulation: driving innovation*. Winter Simulation Conference. 2003, pp. 144–149.

[Twe75]   Richard L. Tweedie. "Sufficient conditions for regularity, recurrence and ergodicity of Markov processes." In: *Mathematical Proceedings of the Cambridge Philosophical Society* 78.1 (1975), pp. 125–136.

[UW09]   Mukhtar Ullah and Olaf Wolkenhauer. "Stochastic Approaches for Systems Biology." In: *Wiley interdisciplinary reviews. Systems biology and medicine* 2 (July 2009), pp. 385–97.

[UW11]   Mukhtar Ullah and Olaf Wolkenhauer. *Stochastic approaches for systems biology*. Springer, 2011.

[VK92]   Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*. Vol. 1. Elsevier, 1992.

[Van10]   Lieven Vandenberghe. *The CVXOPT linear and quadratic cone program solvers*. 2010. URL: http://cvxopt.org/documentation/coneprog.pdf.

[Vir+20]   Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." In: *Nature Methods* 17 (2020), pp. 261–272.

[Vor92]   Henk A Van der Vorst. "Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems." In: *SIAM Journal on scientific and Statistical Computing* 13.2 (1992), pp. 631–644.

[WK19]   Christian Wildner and Heinz Koeppl. "Moment-based variational inference for Markov jump processes." In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6766–6775.

[Wil18]   Darren J Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2018.

[Wil84]   James R Wilson. "Variance reduction techniques for digital simulation." In: *American Journal of Mathematical and Management Sciences* 4.3-4 (1984), pp. 277–312.

[ZK06] IS Zapreev and J-P Katoen. "Safe on-the-fly steady-state detection for time-bounded reachability." In: *Third International Conference on the Quantitative Evaluation of Systems-(QEST'06)*. IEEE. 2006, pp. 301–310.

LIST OF FIGURES

LIST OF TABLES

LIST OF MODELS

ACRONYMS

MPM   Markovian population model

CTMC continuous-time Markov chain

DTMC discrete-time Markov chain

SSA   stochastic simulation algorithm

CME   chemical master equation

SDP   semi-definite program

FPT   first-passage time

MFPT mean first-passage time

ODE   ordinary differential equation

CSL   concurrent separation logic

MTL   metric temporal logic

DNA   deoxyribonucleic acid

CV   control variate

FSP   finite state projection

SEIR   susceptible-exposed-infected-removed

HMM hidden Markov model

IVP   initial value problem