


Bounding the stationary distributions of the chemical master equation via mathematical programming

Cite as: J. Chem. Phys. **151**, 034109 (2019); <https://doi.org/10.1063/1.5100670>

Submitted: 19 April 2019 . Accepted: 24 June 2019 . Published Online: 18 July 2019

Juan Kuntz , Philipp Thomas , Guy-Bart Stan , and Mauricio Barahona 



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[The generalized Boltzmann distribution is the only distribution in which the Gibbs-Shannon entropy equals the thermodynamic entropy](#)

The Journal of Chemical Physics **151**, 034113 (2019); <https://doi.org/10.1063/1.5111333>

[Generating the conformational properties of a polymer by the restricted Boltzmann machine](#)

The Journal of Chemical Physics **151**, 031101 (2019); <https://doi.org/10.1063/1.5103210>

[Large deviations and dynamical phase transitions in stochastic chemical networks](#)

The Journal of Chemical Physics **151**, 064117 (2019); <https://doi.org/10.1063/1.5111110>

Lock-in Amplifiers
... and more, from DC to 600 MHz



Bounding the stationary distributions of the chemical master equation via mathematical programming

Cite as: J. Chem. Phys. 151, 034109 (2019); doi: 10.1063/1.5100670

Submitted: 19 April 2019 • Accepted: 24 June 2019 •

Published Online: 18 July 2019



Juan Kuntz,^{1,2,a)} Philipp Thomas,^{1,b)} Guy-Bart Stan,^{2,c)} and Mauricio Barahona^{1,c)}

AFFILIATIONS

¹Department of Mathematics, Imperial College London, London, United Kingdom

²Department of Bioengineering, Imperial College London, London, United Kingdom

^{a)}Electronic mail: juan.kuntz08@imperial.ac.uk

^{b)}Electronic mail: p.thomas@imperial.ac.uk

^{c)}Authors to whom correspondence should be addressed: g.stan@imperial.ac.uk and m.barahona@imperial.ac.uk

ABSTRACT

The stochastic dynamics of biochemical networks are usually modeled with the chemical master equation (CME). The stationary distributions of CMEs are seldom solvable analytically, and numerical methods typically produce estimates with uncontrolled errors. Here, we introduce mathematical programming approaches that yield approximations of these distributions with computable error bounds which enable the verification of their accuracy. First, we use semidefinite programming to compute increasingly tighter upper and lower bounds on the moments of the stationary distributions for networks with rational propensities. Second, we use these moment bounds to formulate linear programs that yield convergent upper and lower bounds on the stationary distributions themselves, their marginals, and stationary averages. The bounds obtained also provide a computational test for the uniqueness of the distribution. In the unique case, the bounds form an approximation of the stationary distribution with a computable bound on its error. In the nonunique case, our approach yields converging approximations of the ergodic distributions. We illustrate our methodology through several biochemical examples taken from the literature: Schlögl's model for a chemical bifurcation, a two-dimensional toggle switch, a model for bursty gene expression, and a dimerization model with multiple stationary distributions.

© 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5100670>

I. INTRODUCTION

Cell-to-cell variability is pervasive in cell biology. A fundamental source of this variability is the fact that biochemical reactions inside cells often involve only a few molecules per cell.^{1–3} Such reactions are key components in gene regulatory and signaling networks involved in cellular adaptation and cell fate decisions.^{4–6} Mathematically, stochastic reaction networks are modeled using continuous-time Markov chains whose distributions satisfy the chemical master equation (CME). As the availability of accurate single cell measurements widens, it is crucial to develop reliable methods for the analysis of the CME that facilitate parameter inference,^{7,8} both for the identification of molecular mechanisms⁹ and for the design of synthetic cellular circuits.^{10–14}

Significant effort has been devoted to investigating the stationary solutions of CMEs, which determine the long time behavior of the stochastic process.¹⁵ While exact^{16,17} Monte Carlo methods have been developed to sample from stationary solutions of some CMEs, analytical solutions are known only in a few special cases. In general, the CME is considered intractable¹⁸ because, aside of systems with finite state space, it consists of an infinite set of coupled equations.

An approach to circumvent the intractability of the full CME is to compute moments of its stationary solutions. However, moment computations are only exact for networks of unimolecular reactions; in all other cases, the equations of lower moments involve higher moments, leading to an infinite system of coupled equations that cannot be solved analytically. Moment closure schemes,

usually requiring assumptions about the unknown solution, are thus employed to approximate the moments.^{19–23} Yet few of these methods provide quantified approximation errors.²⁰ Another approach is provided by mathematical programming techniques, which have been employed to compute bounds on the moments of Markov processes in various contexts. In such schemes, the finite set of moment equations is supplemented by moment inequalities and the moments are bounded by solving linear programs (LPs)^{24,25} or semidefinite programs (SDPs).^{26,27} Alternatively, when the CME has a unique stationary solution, several truncation-based schemes^{28–32} have been proposed to approximate the solution, although in most cases they do not provide estimates of the error they introduce.

Here, we present two different mathematical programming approaches that yield bounds on, and approximations of, the stationary solutions of the CME. Our first approach builds on our previous work^{27,33,34} and uses semidefinite programming to obtain upper and lower bounds on the moments of stationary solutions of networks with polynomial and rational propensities. The scheme constrains the possible solutions of a truncated, underdetermined set of moment equations by appending semidefinite inequalities that are satisfied by all probability distributions on the state space. Independently of this work, similar approaches have been recently proposed for networks with polynomial propensities.^{33,35–37} Here, we extend the mathematical framework to rational networks of interest in biochemistry, and we state rigorous, checkable mathematical conditions for the validity of the approach.

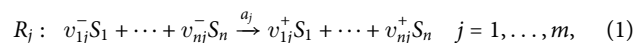
The second approach uses linear programming and a moment bound obtained with our first (SDP) approach to compute lower and upper bounds on stationary averages. In the case of a unique stationary solution, we prove that the LP bounds converge to the true average as the truncation approaches the entire state space. Because stationary averages can be tailored to bound *distributions*, the LP bounds provide approximations with computable errors that converge in total variation to the stationary distribution and its marginals (see also Ref. 38). Additionally, the LP bounds provide a computational test for the uniqueness of the stationary solution, a prerequisite for most other approximation schemes. In the nonunique case, the scheme provides converging approximations of the ergodic distributions.

The paper is organized as follows. In Sec. II, we introduce definitions regarding stationary solutions of the CME. Section III presents the SDP method to bound moments: first, the conceptual framework is introduced analytically for a simple birth-death process that displays a chemical bifurcation, followed by the general computational approach for multispecies networks with rational propensities using semidefinite programming. Section IV presents the LP approach to bound and approximate entire stationary solutions, their averages, and marginals: first, the mathematical framework is introduced through semi-analytical expressions for birth-death processes, followed by the general computational approach for rational networks using linear programming. In Sec. V, we apply the methods to three additional examples: a toggle switch, a model of bursty gene expression with negative feedback, and a model with multiple stationary solutions. We conclude with a discussion in Sec. VI. For completeness, Appendix A presents theoretical results linking stationary distributions of continuous-time chains and the stationary solutions of CMEs, and Appendix B presents a Foster-Lyapunov criterion that guarantees existence and finiteness

of moments of the stationary distribution for the examples in the paper.

II. NOTATION AND DEFINITIONS

Stochastic biochemical kinetics under well-mixed conditions are usually described by a set of m reactions R_j involving n species S_1, S_2, \dots, S_n ,



where $v_{ij}^\pm \in \mathbb{N}$ denote the stoichiometric coefficients and $a_j: \mathbb{N}^n \rightarrow [0, \infty)$ is the propensity of reaction R_j .

Formally, the state of the system is described by the random variable $X(t) = (X_1(t), \dots, X_n(t)) \in \mathbb{N}^n$, a vector with components representing the number of molecules of each species at time t . The dynamical process is modeled with a minimal continuous-time Markov chain³⁹ with rate matrix $Q = (q(x, y))$ defined by

$$q(x, y) := \sum_{j=1}^m a_j(x) (1_{x+v_j}(y) - 1_x(y)), \quad (2)$$

where $v_j := (v_{1j}^+ - v_{1j}^-, \dots, v_{nj}^+ - v_{nj}^-)$ denotes the stoichiometric vector containing the net changes in molecule numbers produced by reaction R_j and 1_y denotes the indicator function of state y ,

$$1_y(x) := \begin{cases} 1, & \text{if } x = y \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The state of the system takes values in a subset $\mathcal{S} \subseteq \mathbb{N}^n$, known as the *state space*, with (possibly infinite) cardinality $|\mathcal{S}|$. The set \mathcal{S} must be chosen such that

$$q(x, y) \geq 0, \quad \forall x \neq y, \quad (4)$$

$$-q(x, x) = \sum_{y \in \mathcal{S}, y \neq x} q(x, y) < \infty, \quad \forall x \in \mathcal{S}, \quad (5)$$

in which case Q is said to be *totally stable* and *conservative*.

If the Markov chain cannot leave the state space in finite time, matrix Q is said to be *regular* (see Appendix A). In this case, the collection of probabilities $p_t(x)$ of observing the chain in state x at time $t \geq 0$ is the only solution of the *chemical master equation* (CME),

$$\frac{dp_t(x)}{dt} = p_t Q(x), \quad p_0(x) = \lambda(x), \quad \forall x \in \mathcal{S}, \quad (6)$$

where we define the vector $p_t := (p_t(x))_{x \in \mathcal{S}}$ and

$$p_t Q(x) := \sum_{y \in \mathcal{S}} p_t(y) q(y, x). \quad (7)$$

Throughout, probability distributions and measures are defined as row vectors.

Any probability distribution $\pi := (\pi(x))_{x \in \mathcal{S}}$ that solves the equation

$$\pi Q(x) = 0, \quad \forall x \in \mathcal{S}, \quad (8)$$

is called a *stationary solution* of the CME. By definition, each stationary solution belongs to the space

$$\ell^1 := \left\{ (\rho(x))_{x \in \mathcal{S}} : \sum_{x \in \mathcal{S}} |\rho(x)| < \infty \right\}$$

of absolutely summable real sequences indexed by states in \mathcal{S} . The set of all stationary solutions forms a convex polytope in ℓ^1 ,

$$\mathcal{P} := \left\{ \pi \in \ell^1 : \begin{aligned} \pi Q(x) &= 0, \quad \forall x \in \mathcal{S}, \\ \pi(\mathcal{S}) &:= \sum_{x \in \mathcal{S}} \pi(x) = 1, \\ \pi &\geq 0 \end{aligned} \right\}, \quad (9)$$

where $\pi \geq 0$ is shorthand for $\pi(x) \geq 0, \forall x \in \mathcal{S}$. For most networks of interest, the stationary solutions determine the long-term behavior of the chain (see [Appendix A](#)). In many cases, we will be interested in obtaining the π -average of a real-valued function f on \mathcal{S} ,

$$\langle f \rangle_\pi := \sum_{x \in \mathcal{S}} f(x) \pi(x).$$

For example, the k th stationary moment of the i th species is $\langle x_i^k \rangle_\pi$.

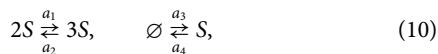
III. BOUNDING THE STATIONARY MOMENTS OF THE CME

As a first use of optimization techniques, we present a systematic approach that yields bounds of increasing tightness on the stationary moments of reaction networks with polynomial or rational propensities, and we give rigorous sufficient conditions for its validity. The moment bounds obtained in this section will be used in conjunction with linear programming to bound the full stationary solutions of the CME in [Sec. IV](#).

To motivate our optimization approach, we first present the mathematical formulation through a simple example, for which explicit analytical expressions can be obtained ([Sec. III A](#)). For more complex systems, the approach can be implemented computationally in a systematic manner through a general semidefinite programming method ([Sec. III B](#)). Readers interested in the computational approach (and not the theory behind it) should skip [Sec. III A](#) and go directly to [Sec. III B](#).

A. A simple analytical example: Moment bounds for Schlögl's model

To illustrate the mathematical framework, consider the classic autocatalytic network with a single species S proposed by Schlögl⁴⁰ as a model for a chemical bifurcation,



with mass-action propensities,

$$a_1(x) := k_1 x(x-1), \quad a_2(x) := k_2 x(x-1)(x-2), \quad (11)$$

$$a_3(x) := k_3, \quad a_4(x) := k_4 x, \quad (12)$$

where $k_1, k_2, k_3, k_4 > 0$ are rate constants. Here, $n = 1$ and the state space is $\mathcal{S} = \mathbb{N}$.

The CME of this network has a unique stationary solution π and all of its moments are finite (see [Appendix B](#)). The reaction network, as encoded in the rate matrix Q , imposes certain relationships between the stationary moments $\langle x^k \rangle_\pi$. Such relations form an infinite system of coupled *stationary moment equations*,

$$\langle Qx^k \rangle_\pi := \sum_{x' \in \mathcal{S}} \sum_{x \in \mathcal{S}} q(x', x) x^k \pi(x') = 0, \quad k \in \mathbb{N}. \quad (13)$$

Except in particular instances, it is not possible to solve this coupled system exactly, and closure approximations are usually adopted by neglecting higher order moments.

An alternative approach is that of mathematical programming, which includes additional constraints in the form of inequalities that must be fulfilled by the moments of distributions. Including such inequalities allows us to obtain feasible regions for the solutions of the system and hence rigorous bounds for the moments. Increasing the number of inequalities considered restricts the feasible region further and makes the bounds tighter.

Let us consider the first moment equation for [\(10\)](#),

$$\langle Qx \rangle_\pi = b_1 \langle 1 \rangle_\pi - b_2 \langle x \rangle_\pi + b_3 \langle x^2 \rangle_\pi - b_4 \langle x^3 \rangle_\pi = 0, \quad (14)$$

where $b_1 := k_3$, $b_2 := k_1 + 2k_2 + k_4$, $b_3 := k_1 + 3k_2$, and $b_4 := k_2$ are positive numbers. Even after noting that

$$\langle 1 \rangle_\pi = 1, \quad (15)$$

as π is a probability distribution, [Eq. \(14\)](#) is underdetermined. However, further relationships between moments can be added to constrain the system. For example, the non-negativity of the variance implies the inequality,

$$\langle x^2 \rangle_\pi - \langle x \rangle_\pi^2 \geq 0. \quad (16)$$

Further inequalities involving higher moments can be built systematically from polynomial functions as follows.

Consider the polynomial $f(x) := f_0 + f_1 x$, where $\mathbf{f} := (f_0, f_1)^T \in \mathbb{R}^2$ is a (column) vector of polynomial coefficients. Clearly, $f^2(x)$ and $x f^2(x)$ are non-negative on $x \in [0, \infty)$. Hence, it follows that

$$\langle f^2 \rangle_\pi = f_0^2 \langle 1 \rangle_\pi + 2f_0 f_1 \langle x \rangle_\pi + f_1^2 \langle x^2 \rangle_\pi \geq 0, \quad (17)$$

$$\langle x f^2 \rangle_\pi = f_0^2 \langle x \rangle_\pi + 2f_0 f_1 \langle x^2 \rangle_\pi + f_1^2 \langle x^3 \rangle_\pi \geq 0. \quad (18)$$

Let us define the vector of moments,

$$z := (\langle 1 \rangle_\pi, \langle x \rangle_\pi, \langle x^2 \rangle_\pi, \langle x^3 \rangle_\pi)^T \in \mathbb{R}^4.$$

Then, the inequalities [\(17\)](#) and [\(18\)](#) are written compactly as

$$\langle f^2 \rangle_\pi = \mathbf{f}^T M_3^0(z) \mathbf{f} \geq 0, \quad (19)$$

$$\langle x f^2 \rangle_\pi = \mathbf{f}^T M_3^1(z) \mathbf{f} \geq 0, \quad (20)$$

where the matrices $M_3^0(y)$ and $M_3^1(y)$ are defined by

$$M_3^0(y) := \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix}, \quad M_3^1(y) := \begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \end{bmatrix},$$

for any four-dimensional vector $y = (y_0, y_1, y_2, y_3)^T \in \mathbb{R}^4$. Since [\(19\)](#) and [\(20\)](#) hold for all $\mathbf{f} \in \mathbb{R}^2$, we have that $M_3^0(z)$ and $M_3^1(z)$ are positive semidefinite (p.s.d.),

$$M_3^0(z) \geq 0, \quad M_3^1(z) \geq 0. \quad (21)$$

From Sylvester's criterion, (21) is equivalent to

$$z_0 \geq 0, \quad z_1 \geq 0, \quad z_2 \geq 0, \quad z_3 \geq 0, \quad (22)$$

$$z_0 z_2 - z_1^2 \geq 0, \quad (23)$$

$$z_1 z_3 - z_2^2 \geq 0. \quad (24)$$

Hence, (23) recovers (16), whereas (24) gives an additional condition involving the first three moments.

Putting (14)–(21) together, we conclude that the singleton set $\{z\}$ of vectors whose entries are composed of the first four moments of the stationary solutions of Schlögl's model belongs to the set

$$\mathcal{E}^3 = \left\{ y \in \mathbb{R}^4 : \begin{cases} y_0 = 1 \\ b_1 y_0 - b_2 y_1 + b_3 y_2 - b_4 y_3 = 0 \\ M_3^0(y) \geq 0 \\ M_3^1(y) \geq 0 \end{cases} \right\}. \quad (25)$$

We say that \mathcal{E}^3 is an *outer approximation* of the set of stationary moment vectors. Hence, the stationary moments have the following lower and upper bounds:

$$L_\alpha^3 := \inf\{y_\alpha : y \in \mathcal{E}^3\} \leq \langle x^\alpha \rangle_\pi \leq \sup\{y_\alpha : y \in \mathcal{E}^3\} =: U_\alpha^3 \\ \alpha = 0, 1, 2, 3. \quad (26)$$

Such bounds are usually handled computationally, but it is illustrative to obtain explicit expressions in this simple case. Combine (24) with the equalities in (25) to get

$$b_1 y_1 + b_3 y_1 y_2 - b_4 y_2^2 - b_2 y_1^2 \geq 0.$$

Assuming $b_3 \geq 2\sqrt{b_2 b_4}$ (the other case is analogous), we use the quadratic formula to obtain the bound

$$r_2^-(y_1) \leq y_2 \leq r_2^+(y_1), \quad (27)$$

$$r_2^\pm(x) := \frac{b_3 x \pm \sqrt{4b_1 b_4 x + (b_3^2 - 4b_2 b_4)x^2}}{2b_4}. \quad (28)$$

Hence, \mathcal{E}^3 can be rewritten equivalently as

$$\mathcal{E}^3 = \left\{ y \in \mathbb{R}^4 : \begin{cases} y_0 = 1 \\ y_1 \geq 0 \\ \max\{y_1^2, r_2^-(y_1)\} \leq y_2 \leq r_2^+(y_1) \\ y_3 = (b_1 - b_2 y_1 + b_3 y_2)/b_4 \end{cases} \right\}. \quad (29)$$

Figure 1(a) shows the projection of \mathcal{E}^3 onto the y_1 – y_2 plane. From (29), it is clear that $(1, 0, 0, b_1/b_4) \in \mathcal{E}^3$, and the lower bounds in (26) for the first two moments are trivial: $(L_1^3, L_2^3) = (0, 0)$. The upper bounds, however, are not. Since $r_2^+(x) > r_2^-(x)$ for all $x > 0$, the upper bounds in (26) are obtained by the northeasternmost intersection of y_1^2 and $r_2^+(y_1)$: $(U_1^3, U_2^3) = (r_4, r_2^+(r_4))$, where r_4 is the rightmost root of $x(b_1 - b_2 x + b_3 x^2 - b_4 x^3) = 0$. In summary, we get

$$0 \leq \langle x \rangle_\pi \leq r_4, \quad 0 \leq \langle x^2 \rangle_\pi \leq r_2^+(r_4). \quad (30)$$

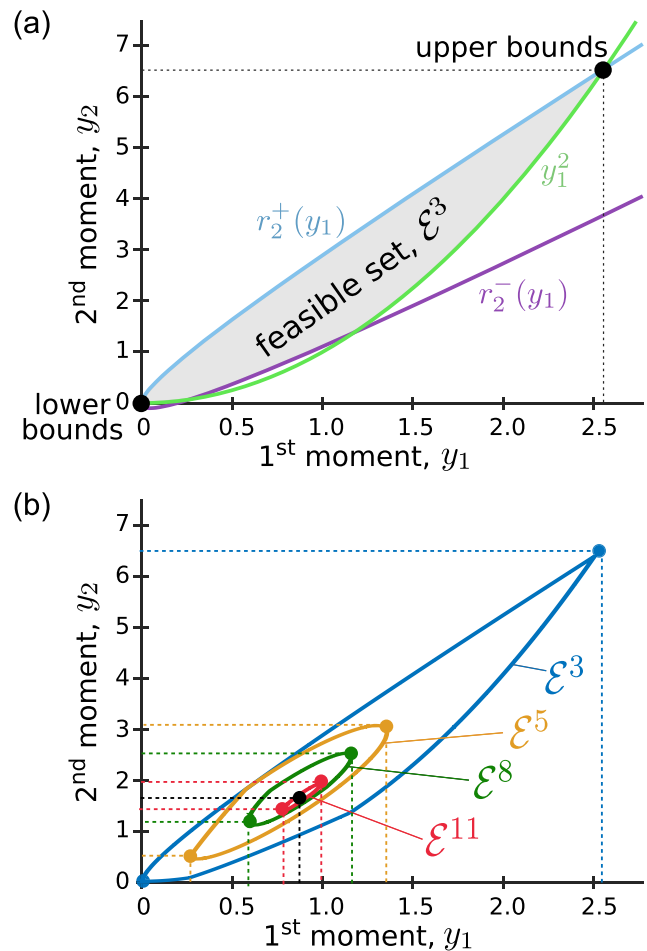


FIG. 1. Outer approximations and bounds for the moments of the stationary solution of Schlögl's model (10). (a) Gray area: projection on the y_1 – y_2 plane of \mathcal{E}^3 (29), an outer approximation of the set of stationary moment vectors. Black dots mark the upper and lower bounds on the first and second moments (30). (b) By appending further moment equations and inequalities, the outer approximations \mathcal{E}^d (45) can be tightened systematically as we increase the order d . The boundaries of the sets \mathcal{E}^d (lines in different colors) were computed explicitly by applying Mathematica's Reduce function to (45). The singleton set of stationary moment vectors (black dot) is always contained in \mathcal{E}^d . The increasingly tighter lower and upper bounds on the moments (49) (colored dots) are computed by solving the SDPs (47) and (48). Parameter values: $k_1 = 1$, $k_2 = 1$, $k_3 = 0.8$, and $k_4 = 1$.

As seen in Fig. 1(a), the analytical bounds based on \mathcal{E}^3 are rough. However, we show in Sec. III B how to obtain tighter bounds systematically by appending further moment equations and inequalities and solving the associated optimizations over higher order sets \mathcal{E}^d .

B. The general approach: Bounding the moments of rational CMEs by solving semidefinite programs

The approach in Sec. III A can be applied to any reaction network (1) with n species and state space \mathcal{S} , as long as the propensities

of its m reactions are rational (or polynomial) functions, i.e., they can be rewritten as

$$a_j(x) := \frac{b_j(x)}{s(x)} \quad j = 1, \dots, m, \quad (31)$$

where b_1, \dots, b_m and s are polynomials on \mathbb{R}^n and the common denominator s satisfies $s(x) > 0, \forall x \in \mathcal{S}$.

To deal with multiple species, we use standard multi-index notation: $x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$, where α is the multi-index $(\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{N}^n$ and $|\alpha| := \alpha_1 + \alpha_2 + \dots + \alpha_n$ is the degree of the monomial. Polynomial functions are expressed in terms of such monomials. For instance, the denominator of (31) is

$$s(x) = \sum_{|\beta| \leq d_s} s_\beta x^\beta, \quad (32)$$

where d_s is the degree of s and we define the (column) vector of coefficients

$$\mathbf{s} := (s_\beta)_{|\beta| \leq d_s}. \quad (33)$$

We also define the stationary rational moments,

$$z_\beta := \left\langle \frac{x^\beta}{s} \right\rangle_\pi, \quad (34)$$

which are directly related to the raw moments,

$$\langle x^\alpha \rangle_\pi = \sum_{|\beta| \leq d_s} s_\beta z_{\alpha+\beta}. \quad (35)$$

The following checkable assumption is a sufficient condition for our general SDP approach to apply to generic reaction networks with rational propensities.

Assumption 1 (Order of the approximation and finiteness of moments). Recall that d_s is the degree of the denominator s in (31). Let us denote the order of the approximation by an integer $d \geq d_s$, and let us compile the stationary rational moments (34) up to order d into the (column) vector,

$$\mathbf{z} := (z_\beta)_{|\beta| \leq d}, \quad (36)$$

of dimension $\#_d := \binom{n+d}{n}$.

We assume that all stationary solutions of the CME have finite rational moments up to order $d + 1$, i.e.,

$$z_\beta = \left\langle \frac{x^\beta}{s} \right\rangle_\pi < \infty \quad \forall \beta : |\beta| \leq d + 1, \quad \forall \pi \in \mathcal{P}.$$

This requirement can be verified using a Foster-Lyapunov criterion as detailed in Appendix B.

Remark 2. From (35) and Assumption 1, it follows that raw moments $\langle x^\alpha \rangle_\pi$ with $|\alpha| \leq d + 1 - d_s$ are also finite.

In order to write down each α -moment, it is helpful to define the associated polynomial function $g_\alpha(x)$,

$$\begin{aligned} s(x) Qx^\alpha &= \sum_{j=1}^m b_j(x) ((x + v_j)^\alpha - x^\alpha) \\ &= \sum_{|\beta| \leq d_{g_\alpha}} (g_\alpha)_\beta x^\beta =: g_\alpha(x), \end{aligned} \quad (37)$$

with degree $d_{g_\alpha} = |\alpha| + d_b - 1$, where $d_b := \max\{d_{b_j}\}$ is the maximum degree of the numerators in (31). We also define the (column) vector of polynomial coefficients,

$$\mathbf{g}_\alpha := \left((g_\alpha)_\beta \right)_{|\beta| \leq d}, \quad (38)$$

where $(g_\alpha)_\beta = 0$ if $|\beta| > d_{g_\alpha}$.

The finiteness of moments guarantees that a subset of moment equations will hold, as stated in the following lemma.

Lemma 3 (The moment equations). If Assumption 1 is satisfied and $\pi \in \mathcal{P}$, then the α -moment equation

$$\langle Qx^\alpha \rangle_\pi = z^T \mathbf{g}_\alpha = 0 \quad (39)$$

holds for every $\alpha \in \mathbb{N}^n$ such that $|\alpha| \leq d - d_b + 1$.

Proof. Consider the adjoint of (8),

$$\langle Qf \rangle_\pi = \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} q(x, y) f(y) \pi(x) = 0, \quad (40)$$

which, by Fubini's theorem, is valid for any f as long as $\sum_{x \in \mathcal{S}} |q(x, x) f(x)| \pi(x)$ is finite.⁴¹ Because

$$s(x) |q(x, x) x^\alpha| = \sum_{j=1}^m b_j(x) x^\alpha = \sum_{j=1}^m \sum_{|\beta| \leq d_b} (b_j)_\beta x^{\alpha+\beta},$$

Assumption 1 implies

$$\sum_{x \in \mathcal{S}} |q(x, x) x^\alpha| \pi(x) \leq \sum_{|\beta| \leq d_b} \left(\sum_{j=1}^m |(b_j)_\beta| \right) z_{\beta+\alpha} < \infty,$$

for all $|\alpha| \leq d - d_b + 1$. Setting $f(x) := x^\alpha$ in (40), we get

$$0 = \langle Qx^\alpha \rangle_\pi = \left\langle \frac{g_\alpha}{s} \right\rangle_\pi = \sum_{|\beta| \leq d} (g_\alpha)_\beta \left\langle \frac{x^\beta}{s} \right\rangle_\pi = z^T \mathbf{g}_\alpha. \quad \square$$

In addition to the moment equations (39), the moments \mathbf{z} satisfy additional constraints. First, since π is a probability distribution, we have

$$z^T \mathbf{s} = \sum_{|\beta| \leq d} s_\beta \left\langle \frac{x^\beta}{s} \right\rangle_\pi = \left\langle \frac{s}{s} \right\rangle_\pi = \langle 1 \rangle_\pi = 1. \quad (41)$$

Furthermore, the rational moments satisfy well-known semidefinite inequalities.^{34,42,43} Specifically, the localizing matrices are positive semidefinite,

$$M_d^i(\mathbf{z}) \geq 0 \quad \forall i = 0, \dots, n, \quad (42)$$

where the $M_d^i(y)$ are defined by

$$\begin{aligned} [M_d^0(y)]_{\alpha\beta} &:= y_{\alpha+\beta}, & \forall \alpha, \beta : |\alpha|, |\beta| \leq \lfloor d/2 \rfloor, \\ [M_d^i(y)]_{\alpha\beta} &:= y_{\alpha+\beta+e_i}, & \forall \alpha, \beta : |\alpha|, |\beta| \leq \lfloor (d-1)/2 \rfloor, \end{aligned}$$

with e_i denoting the i th unit vector and $y \in \mathbb{R}^{\#_d}$.

The inequalities (42) follow from the fact that for any polynomial function $f(x)$ of degree $\lfloor d/2 \rfloor$ with (column) vector of coefficients $\mathbf{f} = (f_\beta)_{|\beta| \leq \lfloor d/2 \rfloor}$, we have

$$\begin{aligned} \mathbf{f}^T M_d^0(z) \mathbf{f} &= \sum_{|\alpha| \leq \lfloor d/2 \rfloor} \sum_{|\beta| \leq \lfloor d/2 \rfloor} f_\alpha f_\beta z_{\alpha+\beta} \\ &= \left\langle \frac{(\sum_{|\alpha| \leq \lfloor d/2 \rfloor} f_\alpha x^\alpha) (\sum_{|\beta| \leq \lfloor d/2 \rfloor} f_\beta x^\beta)}{s} \right\rangle_\pi \\ &= \left\langle \frac{f^2}{s} \right\rangle_\pi \geq 0. \end{aligned} \quad (43)$$

Similarly, it can be shown^{42,43} that

$$\mathbf{f}^T M_d^i(z) \mathbf{f} = \left\langle \frac{x_i f^2}{s} \right\rangle_\pi \geq 0 \quad \forall i = 1, \dots, n. \quad (44)$$

Since (43) and (44) hold for any vector \mathbf{f} , the matrices $M_d^i(z)$ are positive semidefinite.

Remark 4. In the case of Schlögl's model (10), we had $s(x) = 1$, $d_b = 3$, and $\#_d = d + 1$. Equation (14) is the moment equation (39) with $\alpha = 1$, and the matrices in (21) with $d = 3$ are $M_3^0(y)$ and $M_3^1(y)$.

1. Bounding the moments

We can then establish the following lemma regarding outer approximations of the set of rational moments.

Lemma 5 (Outer approximations of the set of rational moments). *If Assumption 1 is satisfied and $\pi \in \mathcal{P}$, the vector of rational moments z belongs to the spectrahedron,*

$$\mathcal{E}^d := \left\{ \begin{array}{l} y^T \mathbf{g}_\alpha = 0 \quad \forall |\alpha| \leq d - d_b + 1, \\ y \in \mathbb{R}^{\#_d} : y^T \mathbf{s} = 1, \\ M_d^i(y) \geq 0 \quad \forall i = 0, 1, \dots, n. \end{array} \right\}, \quad (45)$$

where d_b is the maximum degree of the numerators in (31) and \mathbf{s} , \mathbf{g}_α , and $M_d^i(y)$ are defined in (32), (38), and (42), respectively.

Proof. The proof follows from Lemma 3, (41) and (42). \square

In summary, the vectors of stationary moments of order d are contained in a feasible set \mathcal{E}^d , defined by linear equalities and inequalities, which constitutes an *outer approximation* to the set of moment vectors. There are several implications of this lemma.

First, the outer approximation property implies that extremal points of \mathcal{E}^d provide bounds on the stationary moments. Specifically, the vector in \mathcal{E}^d with largest (respectively, smallest) α -entry provides an upper (respectively, lower) bound on the α -moment. For example, Fig. 1(b) shows the projection of \mathcal{E}^d with increasing d onto the y_1 - y_2 plane for Schlögl's model. The northeasternmost (respectively, southwesternmost) vector of these outer approximations yields upper (respectively, lower) bounds on the first two moments of Schlögl's model.

Second, note that the moment matrix $M_d^i(y)$ is a principal submatrix of $M_{d+1}^i(y)$. Since a matrix is p.s.d. if and only if all of its principal submatrices are p.s.d. and \mathcal{E}^{d+1} includes all moment equations in \mathcal{E}^d , then it follows that every vector in \mathcal{E}^{d+1} (appropriately truncated) belongs to \mathcal{E}^d . As a result, the outer approximations *tighten* around the set of stationary moment vectors with bounds of increasing quality as the order of the approximation d is increased, as seen in Fig. 1(b). These two observations are summarized in the following corollary.

Corollary 6 (Monotonic moment bounds). *Suppose that Assumption 1 is satisfied and $\pi \in \mathcal{P}$. If f is a polynomial of degree $d_f \leq d$, then*

$$L_f^{d_f} \leq L_f^{d_f+1} \leq \dots \leq L_f^d \leq \left\langle \frac{f}{s} \right\rangle_\pi \leq U_f^d \leq \dots \leq U_f^{d_f+1} \leq U_f^{d_f}, \quad (46)$$

where

$$L_f^d := \inf \{ \mathbf{f}^T y : y \in \mathcal{E}^d \}, \quad (47)$$

$$U_f^d := \sup \{ \mathbf{f}^T y : y \in \mathcal{E}^d \}. \quad (48)$$

Proof. Because

$$\left\langle \frac{f}{s} \right\rangle_\pi = \left\langle \frac{\sum_{|\beta| \leq d} f_\beta x^\beta}{s} \right\rangle_\pi = \sum_{|\beta| \leq d} f_\beta \left\langle \frac{x^\beta}{s} \right\rangle_\pi = z^T \mathbf{f},$$

Lemma 5 implies that $L_f^{d_f}, \dots, L_f^d$ and $U_f^{d_f}, \dots, U_f^d$ bound $\langle f/s \rangle_\pi$. As explained in the main text, the monotonicity of the bounds follows from the definition of \mathcal{E}^d and the fact that a matrix is p.s.d. if and only if all of its principal submatrices are p.s.d. and \mathcal{E}^{d+1} includes all moment equations in \mathcal{E}^d . \square

Applying these results to $\langle x^\alpha \rangle_\pi$, the α -moment of the CME, is straightforward. Let $f(x) := s(x)x^\alpha$ and choose $d \geq |\alpha| + d_s$ to obtain the bounds

$$L_\alpha^d := L_f^d, \quad U_\alpha^d := U_f^d. \quad (49)$$

Corollary 6 establishes that outer approximations \mathcal{E}^d of increasing order can be used to compute a *monotonically increasing* (respectively, *decreasing*) sequence of lower (respectively, upper) bounds for $\langle x^\alpha \rangle_\pi$,

$$L_\alpha^{|\alpha|+d_s} \leq \dots \leq L_\alpha^d \leq \langle x^\alpha \rangle_\pi \leq U_\alpha^d \leq \dots \leq U_\alpha^{|\alpha|+d_s}.$$

Remark 7 (The sequence of moment bounds is monotonic but may not converge). *The monotonicity of the bounds does not imply that the gap between the bounds $(U_\alpha^d - L_\alpha^d)$ will converge to zero as $d \rightarrow \infty$. Although in our experience, the bounds often converge numerically, there is no general guarantee for several reasons. First, the stationary solution may not be unique, and in that case, the lower bounds are limited by the stationary solution with the smallest moment, while the upper bounds are limited by that with the largest moment. Even if the solution is unique, the bounds may not converge because the semidefinite conditions are tailored to distributions with support on the non-negative real orthant but not to distributions with support on discrete state spaces,⁴⁴ for which more stringent conditions can be produced at a higher computational cost.^{42,44,45}*

2. Computing moment bounds via semidefinite programming

Given a reaction network with rational propensities and a polynomial f of degree d_f , the moment bounds (47) and (48) are the extreme points of the linear functional $y \mapsto \mathbf{f}^T y$ over the set \mathcal{E}^d , which is defined by linear equalities and semidefinite inequalities. Hence, computing the bounds amounts to solving a *semidefinite program* (SDP), a convex optimization problem for which there exist efficient computational tools. Therefore, instead of *ad hoc* analytical manipulations, such as those leading to (30), a general procedure by constructing and solving the SDPs systematically is implemented as follows:

1. Rewrite the reaction propensities in the form (31) removing all common factors and setting s to be the lowest common denominator.
2. Verify the existence of stationary solutions π and choose the order of the approximation d , an integer $d \geq d_f$ for which the $d + 1$ stationary moments are finite (Assumption 1) using a Foster-Lyapunov criterion (Theorem 31 in Appendix B).

3. Compute the bounds L_f^d and U_f^d by solving the two SDPs (47) and (48). We set up the SDPs using the modeling package YALMIP⁴⁶ and solve them using the multiprecision solver SDPA-GMP⁴⁷ with the interface mpYALMIP.⁴⁸ Examples of computation times are given in the figure captions.

SDPs involving high order moments can be numerically ill-conditioned.^{27,44,49} Although the origin of this numerical instability remains an open problem, our computations suggest that it could be the result of the rapid growth of moments, which leads to ill-conditioned moment matrices. Such disparity is problematic for standard double-precision SDP solvers, but we have mitigated it with the multiprecision solver SDPA-GMP⁴⁷ as in Ref. 49. Alternatively, one can scale the moments^{27,44,49} or adapt recently developed specialized solvers.^{50,51}

4. Evaluate the error of the bounds by computing the gap $U_f^d - L_f^d$. If the gap is unsatisfactorily large, increase the order of the approximation and return to step 2 to compute new bounds.

Corollary 6 guarantees that the bounds will not loosen as we increase d , yet the bounds may stagnate (Remark 7). In this

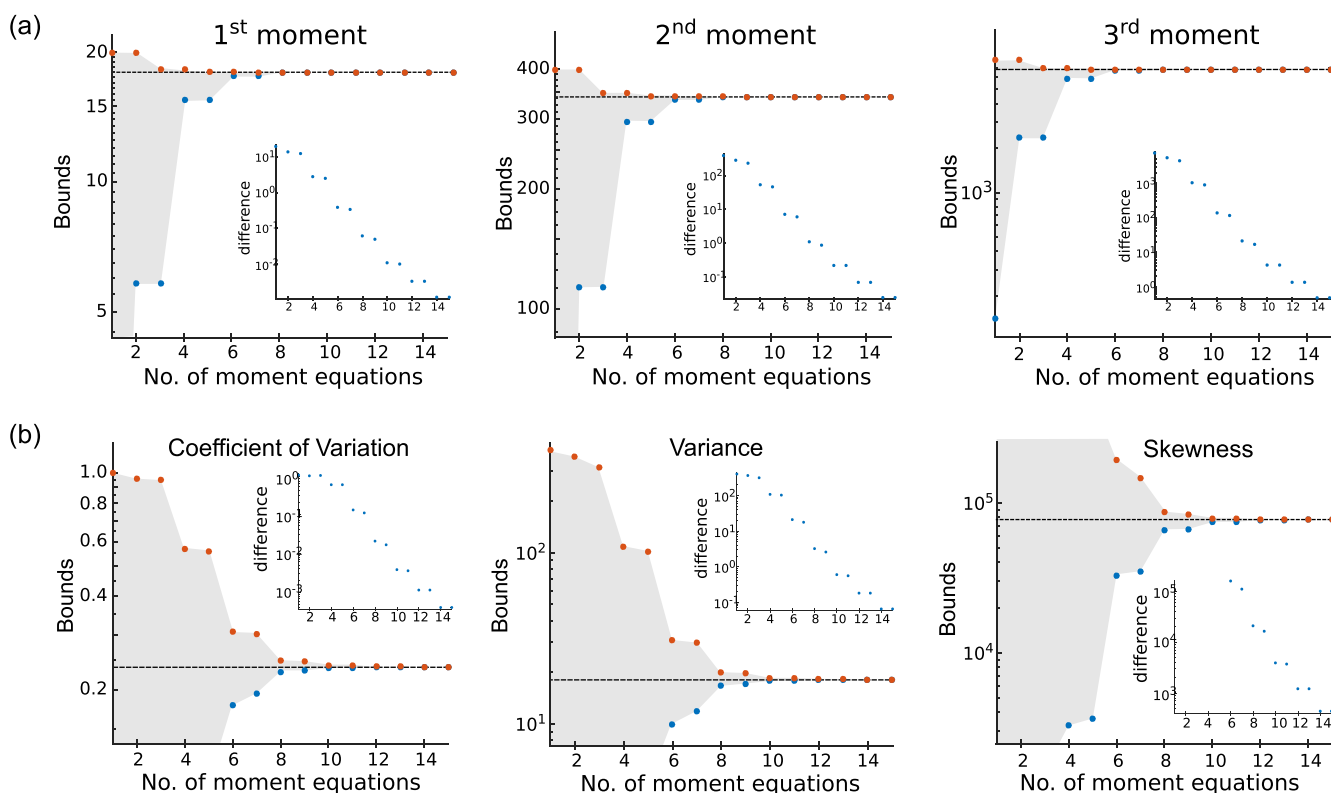


FIG. 2. Moment bounds for Schlögl's model (10) using SDPs. (a) Upper bounds (U_{α}^d , red circles) and lower bounds (L_{α}^d , blue circles) for the first three moments ($\langle x^{\alpha} \rangle_{\pi}$, $\alpha = 1, 2, 3$) computed for increasing order of the approximation, d (no. of moment equations = $d - 2$). The bounds for the first two moments computed by solving the SDPs with $d = 3$ coincide with the analytical expressions (30). The bounds approach the true moments (dashed lines). Inset: the gap between upper and lower bounds decreases to zero as d increases. (b) The moment bounds in (a) are used to obtain bounds of three typical statistics: coefficient of variation, variance, and skewness. Insets: the gap between lower and upper bounds also decreases to zero. For each moment, we computed 30 bounds (upper and lower, $d = 3, \dots, 17$) for a solver time of 29 s (1 s per bound). Parameters: $k_1 = 6$, $k_2 = 1/3$, $k_3 = 50$, and $k_4 = 3$, which correspond to a unimodal stationary solution.

case, we recommend breaking the impasse by employing the LP approach of Sec. IV [see Fig. 7(b)].

As an example of this procedure, Fig. 2(a) shows how the computed upper and lower bounds (L_α^d, U_α^d) for the first three stationary moments of Schlögl's model become tighter as we increase d , the order of the approximation. Figure 2(b) combines the moment bounds to obtain bounds on commonly used statistics, e.g., variance, coefficient of variation, and skewness.

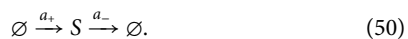
IV. BOUNDING AND APPROXIMATING THE STATIONARY SOLUTIONS OF THE CME

The semidefinite programming scheme in Sec. III B 2 allows us to compute bounds of the stationary moments of the CME by constructing and optimizing over outer approximations of the set of stationary moments. In this section, we go further and introduce a linear programming scheme that yields bounds on the full stationary solutions of the CME, their marginals, and their averages by constructing and optimizing over outer approximations of the set of stationary solutions. To do so, we use a moment bound obtained using the approach in Sec. III B 2. Note that the approximations in this section only involve linear inequalities (instead of semidefinite ones). Hence, the optimizations to be solved are linear programs (instead of SDPs), a simpler subclass of convex optimization problems for which mature, industrial solvers⁵² are available.

As for the SDP scheme above, we introduce the mathematical approach through a simple semi-analytic example (i.e., birth-death processes) in Sec. IV A and then present the computational framework for general CMEs in Sec. IV B. Readers interested in the computational implementation (and not the mathematical background) should skip Sec. IV A and go directly to Sec. IV B.

A. A simple example: Bounding the stationary solution of birth-death processes

Birth-death processes are one-species reaction networks ($n = 1$) with state space $\mathcal{S} = \mathbb{N}$ whose value $x \in \mathbb{N}$ changes by ± 1 in each reaction,



The specific birth-death process is defined by the functional form of the given propensities $a_+(x)$ and $a_-(x)$.

The stationary equations $\pi Q = 0$ then read

$$a_-(1)\pi(1) - a_+(0)\pi(0) = 0, \quad (51)$$

$$a_-(x+1)\pi(x+1) - (a_+(x) + a_-(x))\pi(x) + a_+(x-1)\pi(x-1) = 0, \quad x \geq 1. \quad (52)$$

Assuming nonvanishing death rates

$$a_-(x) > 0, \quad \forall x \geq 1, \quad (53)$$

it is well known that the unique stationary solution is⁵³

$$\pi(x) = \left[\prod_{z=1}^x \frac{a_+(z-1)}{a_-(z)} \right] \pi(0) =: \gamma(x) \pi(0), \quad \forall x \geq 0, \quad (54)$$

with $\pi(0)$ given by the normalization condition,

$$\pi(0) = \frac{1}{\sum_{x=0}^{\infty} \gamma(x)} =: \frac{1}{\gamma(\mathcal{S})}, \quad (55)$$

where we have introduced the notation for sums over sets,

$$\sum_{x \in \mathcal{S}} \gamma(x) =: \gamma(\mathcal{S}). \quad (56)$$

Hence, birth-death processes have at most one stationary solution, which exists if and only if $\gamma(\mathcal{S})$ is finite.

1. Semi-analytical approach for bounds and approximation

For most birth-death processes, no closed-form expression for $\gamma(\mathcal{S})$ is known and, consequently, the stationary solution cannot be computed exactly. However, we can obtain upper and lower bounds for the distribution, as follows.

Let us consider a *state space truncation*

$$\mathcal{S}_r := \{x \in \mathbb{N} : x^\alpha < r\} = \{0, 1, \dots, \lceil r^{1/\alpha} \rceil - 1\}, \quad (57)$$

with size $|\mathcal{S}_r| = \lceil r^{1/\alpha} \rceil$ controlled by the parameters $\alpha \in \mathbb{Z}_+$ and $r > 0$. Let \mathcal{S}_r^c denote its complement, i.e., the set of states outside of the truncation \mathcal{S}_r .

Let us assume that we have available an upper bound on the stationary α -moment,

$$\langle x^\alpha \rangle_\pi \leq c. \quad (58)$$

Note that for rational propensities, such a bound can be computed with the SDP scheme in Sec. III.

Upper bound: An easy upper bound on $\pi(0)$ is obtained by truncating the sum in (55) to get

$$\frac{1}{\sum_{x \in \mathcal{S}_r} \gamma(x)} \geq \frac{1}{\sum_{x=0}^{\infty} \gamma(x)} = \pi(0),$$

whence it follows that

$$\pi(x) \leq \frac{\gamma(x)}{\gamma(\mathcal{S}_r)} =: u_x^r \quad \forall x \in \mathcal{S}_r. \quad (59)$$

Lower bound: Using (58), we obtain a bound on the probability mass m_r outside of the truncation,

$$m_r := \sum_{x \notin \mathcal{S}_r} \pi(x) \leq \frac{1}{r} \sum_{x \notin \mathcal{S}_r} x^\alpha \pi(x) \leq \frac{\langle x^\alpha \rangle_\pi}{r} \leq \frac{c}{r} =: \varepsilon_r. \quad (60)$$

We say that ε_r is a *tail bound*.

A lower bound on $\pi(0)$ then follows from (54)–(60),

$$\pi(0) = \frac{1 - m_r}{\sum_{x \in \mathcal{S}_r} \gamma(x)} \geq \frac{1 - \varepsilon_r}{\gamma(\mathcal{S}_r)}, \quad (61)$$

whence we obtain a lower bound for $\pi(x)$,

$$\pi(x) \geq u_x^r (1 - \varepsilon_r) =: l_x^r, \quad \forall x \in \mathcal{S}_r. \quad (62)$$

Convergent bounds: We have thus shown that

$$l_x^r \leq \pi(x) \leq u_x^r, \quad \forall x \in \mathcal{S}_r, \quad (63)$$

and it is easy to see that both bounds converge to the stationary solution as the size of the truncation grows: as $r \rightarrow \infty$, both $u_x^r \rightarrow \pi(x)$ and $l_x^r \rightarrow \pi(x)$. This follows from (54) and (55) and $\varepsilon_r \rightarrow 0$.

a. *Approximating the distribution and the approximation error.* Motivated by these facts, we define the two following measures (lower and upper bounds padded with zeros):

$$l^r := (l^r(x))_{x \in \mathcal{S}}, \quad l^r(x) := \begin{cases} l_x^r & \text{if } x \in \mathcal{S}_r \\ 0 & \text{if } x \notin \mathcal{S}_r \end{cases}, \quad (64)$$

$$u^r := (u^r(x))_{x \in \mathcal{S}}, \quad u^r(x) := \begin{cases} u_x^r & \text{if } x \in \mathcal{S}_r \\ 0 & \text{if } x \notin \mathcal{S}_r \end{cases}, \quad (65)$$

and introduce them as *approximations* for π ,

$$\tilde{\pi} \simeq \pi, \text{ where } \tilde{\pi} = l^r \text{ or } \tilde{\pi} = u^r.$$

We quantify the *approximation error* of $\tilde{\pi}$ with the total variation norm,

$$\|\pi - \tilde{\pi}\| = \sup_{A \subseteq \mathcal{S}} |\pi(A) - \tilde{\pi}(A)|, \quad (66)$$

where $\pi(A)$ and $\tilde{\pi}(A)$ are sums over sets, defined in (56).

For u^r and l^r , the approximation error can be characterized further. Using (59)–(65), we have

$$\begin{aligned} \|\pi - l^r\| &= \pi(\mathcal{S}) - l^r(\mathcal{S}) = 1 - \sum_{x \in \mathcal{S}_r} l_x^r \\ &= 1 - (1 - \varepsilon_r) \sum_{x \in \mathcal{S}_r} u_x^r = \varepsilon_r, \end{aligned} \quad (67)$$

$$\begin{aligned} \|\pi - u^r\| &= \max\{u^r(\mathcal{S}_r) - \pi(\mathcal{S}_r), \pi(\mathcal{S}_r^c)\} \\ &= \max\{1 - (1 - m_r), m_r\} = m_r \leq \varepsilon_r. \end{aligned} \quad (68)$$

Since $\varepsilon_r = c/r \rightarrow 0$ as $r \rightarrow \infty$, it thus follows that both l^r and u^r converge in total variation to π .

We summarize these findings in the following theorem:

Theorem 8 (Bounds and approximations of the stationary solution of birth-death processes). *Consider any birth-death process (50) with nonvanishing decay rates (53) and finite sum $\gamma(\mathcal{S})$ (55) such that it has a unique stationary solution π (54). Suppose that π satisfies the moment bound (58) and let $\mathcal{S}_r \subseteq \mathcal{S}$ be the truncation (57) of the state space controlled by the parameters $r, \alpha \in \mathbb{Z}_+$, with tail bound $m_r = \pi(\mathcal{S}_r^c) \leq c/r = \varepsilon_r$. Then, the following hold:*

- (i) *The values of the distribution over the truncation are bounded above and below,*

$$l_x^r \leq \pi(x) \leq u_x^r, \quad \forall x \in \mathcal{S}_r,$$

where $u_x^r = \gamma(x)/\gamma(\mathcal{S}_r)$ and $l_x^r = u_x^r(1 - \varepsilon_r)$.

- (ii) *The measures $l^r = (l^r(x))_{x \in \mathcal{S}}$ and $u^r = (u^r(x))_{x \in \mathcal{S}}$ defined in (64) and (65) approximate the solution with approximation errors,*

$$\|\pi - l^r\| = \varepsilon_r \text{ and } \|\pi - u^r\| = m_r.$$

- (iii) *The bounds vary monotonically with r ,*

$$l_x^r \leq l_x^{r+1} \leq \dots \leq \pi(x) \leq \dots \leq u_x^{r+1} \leq u_x^r, \quad \forall x \in \mathcal{S}_r,$$

and the sequences of approximations converge in total variation to π ,

$$\lim_{r \rightarrow \infty} \|\pi - l^r\| = \lim_{r \rightarrow \infty} \|\pi - u^r\| = 0.$$

Proof. This follows from (59)–(68) and Corollary 6. \square

Remark 9. If Assumption 1 holds, then π satisfies the moment bound (58) with $c = U_\alpha^d$, where $\alpha \in \{1, \dots, d - d_s\}$ and U_α^d is defined in (48).

b. *An application of theorem 8: Schlögl's model.* To illustrate our results, we apply Theorem 8 to compute bounds on the unique stationary solution of Schlögl's model (10)–(12). This model is a birth-death process for which an explicit analytical stationary solution can be obtained, thus allowing us to test the results directly without any simulations.

Through some analytical manipulations, the solution of Schlögl's model can be obtained explicitly in terms of

$$\frac{1}{\pi(0)} = {}_2F_2\left(-\frac{c_1 + 1}{2}, \frac{c_1 - 1}{2}; -\frac{c_2 + 1}{2}, \frac{c_2 - 1}{2}; \frac{k_1}{k_2}\right), \quad (69)$$

where ${}_2F_2$ denotes the generalized hypergeometric function; $c_1 := \sqrt{1 - 4k_3/k_1}$; and $c_2 := \sqrt{1 - 4k_4/k_2}$. The stationary solution goes from being unimodal [black line, Fig. 3(a)] to bimodal [black line, Fig. 3(b)] depending on the parameter values, analogously to a bifurcation.

In Figs. 3(a) and 3(b), we compare the approximations l^r and u^r given in Theorem 8 (color shades) to the analytical solution (black lines) of the unimodal and bimodal cases. As the size of the truncation (controlled by the parameter r) is increased, the bounds tighten around the analytical solution. In Figs. 3(c) and 3(d), we show that the approximation error tends to zero as the size of the truncation $|\mathcal{S}_r| = \lceil r^{1/\alpha} \rceil$ is increased. In the unimodal case [Fig. 3(c)], the approximation error decreases rapidly when $|\mathcal{S}_r|$ is larger than the mode. Furthermore, when the truncation is sufficiently large, employing bounds on higher order moments (larger α) provides tighter bounds and smaller approximation errors. On the other hand, if the size of the truncation is smaller than the mode, using higher order moments does not necessarily improve the approximation error. A similar dependence of the approximation error is observed in the bimodal case [Fig. 3(d)], but the approximation error only decreases when the truncation size is larger than the second (larger) mode. This example shows how the ability to compute error bounds can reveal the presence of modes outside of the truncation.

2. Reformulation of the bounds as optimizations

The truncation method leading to Theorem 8 relies on the detailed balanced structure of birth-death processes. However, this semi-analytic method is not generalizable to arbitrary reaction networks. Instead, the bounds can be reformulated as an equivalent (and generalizable) optimization problem, as follows.

Consider a truncation \mathcal{S}_r (57) controlled by the parameters $\alpha, r \in \mathbb{Z}_+$ with tail bound $m_r = \pi(\mathcal{S}_r^c) \leq c/r = \varepsilon_r$.

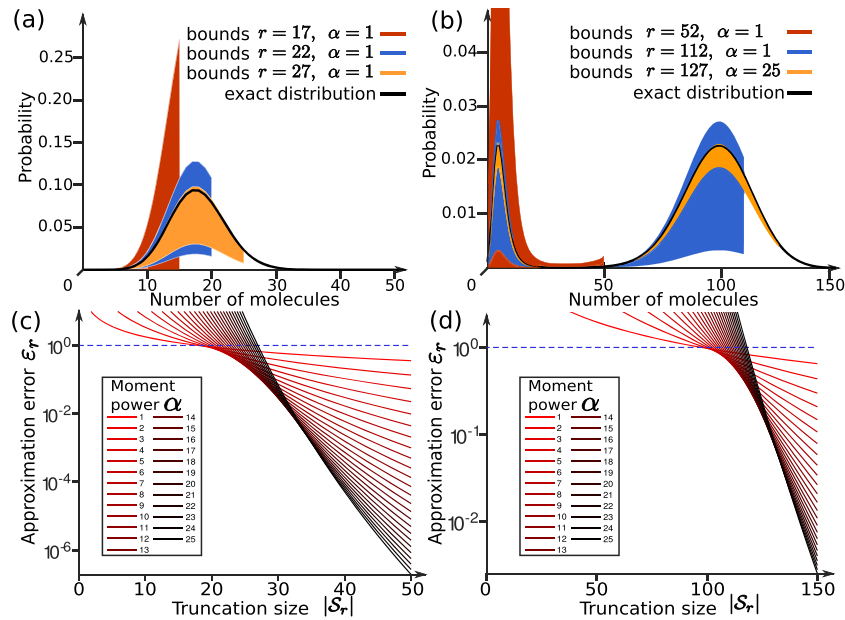


FIG. 3. Bounding the stationary solution of Schlögl's model (10). The stationary solution of (10) can be unimodal (a) or bimodal (b) depending on the parameters. (a) Shadings show the tightening gap (l^r, u^r) between upper and lower bounds on the stationary solution for truncations \mathcal{S}_r of increasing size ($|\mathcal{S}_r| = \lceil r^{1/\alpha} \rceil$). We use $m_r \leq U_\alpha^{25}/r$ with $U_1^{25} = 17.5$ computed as in Sec. III (solver time = 7 s). The exact solution, given by (54) and (69), is shown for comparison (black line). (b) Same as (a) but for the bimodal case, with tail bounds computed using $U_{25}^{25} = 98.0$, $U_{25}^{25} = 6.37 \times 10^{31}$ (solver time = 8 and 7 s, respectively). (c) The approximation error of the lower bound approximation, ε_r , for the unimodal case decreases as $|\mathcal{S}_r|$ increases, shown here for various values of α . (d) Same as (c) but for the bimodal case. Note that only when the truncation includes enough states [different values of $|\mathcal{S}_r|$ for (c) and (d)], does the error fall below the dashed line $\varepsilon_r < 1$ so that the bounds provide information about the stationary solution. For the bimodal case, the error indicates the presence of a second mode outside of the truncation when $|\mathcal{S}_r|$ is too small. Parameters: (a): $k_1 = 6$, $k_2 = 1/3$, $k_3 = 50$, and $k_4 = 3$; (b): $k_1 = 1/9$, $k_2 = 1/1215$, $k_3 = 27/2$, and $k_4 = 59/20$.

Definition 10 (Restriction of π to \mathcal{S}_r). The restriction of π to \mathcal{S}_r is

$$\pi|_r(x) := \begin{cases} \pi(x) & \text{if } x \in \mathcal{S}_r \\ 0 & \text{if } x \notin \mathcal{S}_r \end{cases}, \quad \forall x \in \mathcal{S}. \quad (70)$$

The restriction $\pi|_r$ belongs to the convex polytope,

$$\mathcal{P}^r = \left\{ \begin{array}{l} \pi^r Q(x) = 0, \forall x \in \mathcal{N}_r \\ \pi^r(\mathcal{S}_r^c) = 0 \\ \pi^r \in \ell^1 : \pi^r \geq 0 \\ 1 - \varepsilon_r \leq \pi^r(\mathcal{S}) \leq 1 \\ \langle x^\alpha \rangle_{\pi^r} \leq c \end{array} \right\}, \quad (71)$$

where $\mathcal{N}_r := \{0, 1, \dots, \lceil r^{1/\alpha} \rceil - 1\}$, which follows directly from the definition of the restriction (70), the tail bound (60), and the fact that the stationary equations (51) and (52) with $x < \lceil r^{1/\alpha} \rceil - 1$ only involve states inside of the truncation.

From the definition of the polytope (71), we can show that the bounds of the stationary solution in Theorem 8 (i) are recovered by optimizing over \mathcal{P}^r , as stated in the following theorem.

Theorem 11 (Bounds and LP formulation). The bounds l^r and u^r in (64) and (65) are obtained by optimizing over the polytope \mathcal{P}^r ,

$$l^r(x) = \inf\{\pi^r(x) : \pi^r \in \mathcal{P}^r\}, \\ u^r(x) = \sup\{\pi^r(x) : \pi^r \in \mathcal{P}^r\}, \quad \forall x \in \mathcal{S}.$$

Proof. We only present the argument for the upper bounds—the proof for the lower bounds is analogous. If $x \notin \mathcal{S}_r$, the result is trivial. A distribution π^r satisfies (71) if and only if $\pi^r(x) = \gamma(x)\pi^r(0)$, $\forall x \in \mathcal{S}_r$, where $\gamma(x)$ is given in (54). Therefore, we have

$$\sup\{\pi^r(x) : \pi^r \in \mathcal{P}^r\} = \gamma(x)(\sup\{\pi^r(0) : \pi^r \in \mathcal{P}^r\}) \\ \leq \frac{\gamma(x)}{\gamma(\mathcal{S}_r)} = u^r(x), \quad \forall x \in \mathcal{S}_r, \quad (72)$$

which follows from $\pi^r(0) = \pi^r(\mathcal{S}_r)/\gamma(\mathcal{S}_r) \leq 1/\gamma(\mathcal{S}_r)$. However, u^r clearly satisfies all constraints in (71), including the moment constraint,

$$c \geq \langle x^\alpha \rangle_{\pi} = \frac{\gamma(\mathcal{S}_r)}{\gamma(\mathcal{S})} \sum_{x \in \mathcal{S}_r} \frac{\gamma(x)}{\gamma(\mathcal{S}_r)} x^\alpha + \frac{\gamma(\mathcal{S}_r^c)}{\gamma(\mathcal{S})} \sum_{x \in \mathcal{S}_r^c} \frac{\gamma(x)}{\gamma(\mathcal{S}_r^c)} x^\alpha \\ > \frac{\gamma(\mathcal{S}_r)}{\gamma(\mathcal{S})} \langle x^\alpha \rangle_{u^r} + \frac{\gamma(\mathcal{S}_r^c)}{\gamma(\mathcal{S})} \langle x^\alpha \rangle_{u^r} = \langle x^\alpha \rangle_{u^r},$$

which follows from $\pi(x) = \gamma(x)/\gamma(\mathcal{S})$ and the inequality

$$\sum_{x \in \mathcal{S}_r^c} \frac{\gamma(x)}{\gamma(\mathcal{S}_r^c)} x^\alpha \geq \sum_{x \in \mathcal{S}_r^c} \frac{\gamma(x)}{\gamma(\mathcal{S}_r^c)} r = \sum_{x \in \mathcal{S}_r^c} \frac{\gamma(x)}{\gamma(\mathcal{S}_r)} r \\ > \sum_{x \in \mathcal{S}_r} \frac{\gamma(x)}{\gamma(\mathcal{S}_r)} x^\alpha = \langle x^\alpha \rangle_{u^r}.$$

Hence, $u^r \in \mathcal{P}^r$ and together with (72), this completes the proof. \square

Importantly, the definition (71) only involves linear equations and inequalities. Therefore, optimizing over the polytope consists of solving a *linear program*, a class of optimizations for which there exist powerful computational platforms and algorithms. Furthermore, this optimization reformulation can be extended seamlessly to arbitrary networks, as expanded in Sec. IV B.

B. Generalization to arbitrary networks via linear programming

We now generalize the optimization approach to obtain bounds and approximations with controlled errors of the stationary solutions of arbitrary reaction networks.

Let us consider a reaction network (1) with state space \mathcal{S} , rate matrix Q satisfying (4) and (5), and stationary solutions $\pi \in \ell^1$ that form the polytope \mathcal{P} (9).

To characterize the solutions of the CME, we choose a *normlike* function w , which plays the same role as the moment function (x^α) in Sec. IV A 2.

Definition 12 (Normlike function). A function $w : \mathcal{S} \rightarrow \mathbb{R}$ is *normlike* if it is *non-negative*,

$$w(x) \geq 0, \forall x \in \mathcal{S}, \quad (73)$$

and has *finite sublevel sets*

$$\mathcal{S}_r := \{x \in \mathcal{S} : w(x) < r\}. \quad (74)$$

Furthermore, we require that the growth of w be *dominated* by the stationary solution π so that its expectation with respect to π is finite. We summarize these requirements in the following checkable assumption.

Assumption 13 (Existence of CME solutions and moment bound). We assume that the CME has at least one stationary solution π and that we have available a normlike function w with sublevel sets \mathcal{S}_r such that every stationary solution π satisfies

$$\langle w \rangle_\pi = \sum_{x \in \mathcal{S}} w(x)\pi(x) \leq c, \quad (75)$$

where c is a known constant. This inequality can be thought of as a generalization of (58); hence, we refer to it as a *moment bound*.

The existence of the stationary solutions can be verified on a case by case basis using a Foster-Lyapunov criterion (e.g., Theorem 31 in Appendix B).

Regarding the moment bound, in the case of networks with rational propensities satisfying Assumption 1, w can be chosen to be any normlike rational function with numerator of degree d and the bounding constant c can then be computed using the SDP approach of Sec. III. For general reaction networks, the moment bound can be obtained using Foster-Lyapunov criteria⁴¹ (see Appendix B).

In analogy with Lyapunov theory, the sublevel sets (74) of w play an important role in characterizing the stationary solutions of the CME. Specifically, we use the sublevel sets \mathcal{S}_r as our state space truncations, noting that (75) allows us to establish a bound on the mass of the tail of the distribution outside of \mathcal{S}_r ,

$$m_r := \pi(\mathcal{S}_r^c) \leq \frac{1}{r} \sum_{x \notin \mathcal{S}_r} w(x)\pi(x) \leq \frac{\langle w \rangle_\pi}{r} \leq \frac{c}{r} := \varepsilon_r. \quad (76)$$

Just as in Sec. IV A, this choice yields a sequence of increasing truncations that approach the entire state space,

$$\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \dots, \quad \bigcup_{r=1}^{\infty} \mathcal{S}_r = \mathcal{S}.$$

For each truncation, let \mathcal{N}_r denote the set of states $x \in \mathcal{S}_r$ that cannot be reached in a single jump from outside of the truncation,

$$\mathcal{N}_r := \{x \in \mathcal{S}_r : q(z, x) = 0, \forall z \notin \mathcal{S}_r\}, \quad (77)$$

and the associated convex polytope,

$$\mathcal{P}^r := \left\{ \begin{array}{l} \pi^r Q(x) = 0, \quad \forall x \in \mathcal{N}_r \\ \pi^r(\mathcal{S}_r^c) = 0 \\ \pi^r \in \ell^1 : \pi^r \geq 0 \\ 1 - \varepsilon_r \leq \pi^r(\mathcal{S}) \leq 1 \\ \langle w \rangle_{\pi^r} \leq c \end{array} \right\}, \quad (78)$$

which, analogously to (71), includes all the stationary equations that only involve states in \mathcal{S}_r .

We then have the following lemma:

Lemma 14 (Outer approximations of \mathcal{P}). Suppose that Assumption 13 holds and let $\pi|_{\mathcal{S}_r}$ be the restriction of π to \mathcal{S}_r , as defined in (70). If $\pi \in \mathcal{P}$, then $\pi|_{\mathcal{S}_r} \in \mathcal{P}^r$.

Proof. This follows directly from (76) and the fact that $\pi Q(x) = \pi|_{\mathcal{S}_r} Q(x)$, $\forall x \in \mathcal{N}_r$. \square

The outer approximation property means that optimizing over \mathcal{P}^r provides convergent bounds on the averages of functions f on the state space, as summarized in the following theorem.

Theorem 15 (Convergent bounds of stationary averages). Consider a reaction network (1) with state space \mathcal{S} , rate matrix Q satisfying (2)–(5), and stationary solutions π forming the set \mathcal{P} (9) and suppose that Assumption 13 holds.

If $\pi \in \mathcal{P}$ and $f : \mathcal{S} \rightarrow \mathbb{R}$ is any real-valued function, then we can bound its averages over the restrictions,

$$l_f^r \leq \langle f \rangle_{\pi|_{\mathcal{S}_r}} \leq u_f^r, \quad \forall r \in \mathbb{Z}_+, \quad (79)$$

where $\pi|_{\mathcal{S}_r}$ is the restriction of π to \mathcal{S}_r defined in (70) and the bounds are given by

$$\begin{aligned} l_f^r &:= \inf\{\langle f \rangle_{\pi^r} : \pi^r \in \mathcal{P}^r\}, \\ u_f^r &:= \sup\{\langle f \rangle_{\pi^r} : \pi^r \in \mathcal{P}^r\}. \end{aligned} \quad (80)$$

If we have additional information on f , we have the following bounds on the full π -averages:

- (i) If $f(x) \geq 0, \forall x \notin \mathcal{S}_r$, then $l_f^r \leq \langle f \rangle_\pi, \forall r \in \mathbb{Z}_+$.
- (ii) If $f(x) \leq 0, \forall x \notin \mathcal{S}_r$, then $\langle f \rangle_\pi \leq u_f^r, \forall r \in \mathbb{Z}_+$.
- (iii) If $\langle |f| \rangle_\pi < \infty$ (i.e., f is π -integrable), then

$$l_f^r - c \left(\sup_{x \notin \mathcal{S}_r} \frac{|f(x)|}{w(x)} \right) \leq \langle f \rangle_\pi \leq u_f^r + c \left(\sup_{x \notin \mathcal{S}_r} \frac{|f(x)|}{w(x)} \right). \quad (81)$$

- (iv) If the growth of f is strictly dominated by that of w , i.e.,

$$\lim_{r \rightarrow \infty} \sup_{x \in S_r} \frac{|f(x)|}{w(x)} = 0, \quad (82)$$

then f is π -integrable for all $\pi \in \mathcal{P}$ and the sequences of bounds from below $(l_f^r)_{r \in \mathbb{Z}_+}$ and above $(u_f^r)_{r \in \mathbb{Z}_+}$ converge,

$$\begin{aligned} \lim_{r \rightarrow \infty} l_f^r &= l_f := \inf\{\langle f \rangle_\pi : \pi \in \mathcal{P}\}, \\ \lim_{r \rightarrow \infty} u_f^r &= u_f := \sup\{\langle f \rangle_\pi : \pi \in \mathcal{P}\}. \end{aligned} \quad (83)$$

Proof. Equation (79) follows directly from Lemma 14.

(i) and (ii) follow from (79) and $\langle f \rangle_\pi = \langle f \rangle_{\pi_r} + \sum_{x \notin S_r} f(x)\pi(x)$.

(iii) is a consequence of (79), the moment bound (75), and the following inequality:

$$\begin{aligned} \sum_{x \notin S_r} |f(x)|\pi(x) &\leq \left(\sup_{x \notin S_r} \frac{|f(x)|}{w(x)} \right) \sum_{x \notin S_r} w(x)\pi(x) \\ &\leq \left(\sup_{x \notin S_r} \frac{f(x)}{w(x)} \right) \langle w \rangle_\pi. \end{aligned} \quad (84)$$

(iv) has two parts. First, the π -integrability of f follows from

$$\begin{aligned} \langle |f| \rangle_\pi &= \langle |f| \rangle_{\pi_r} + \sum_{x \notin S_r} |f(x)|\pi(x) \\ &\leq \langle |f| \rangle_{\pi_r} + \left(\sup_{x \notin S_r} \frac{f(x)}{w(x)} \right) \langle w \rangle_\pi < \infty. \end{aligned}$$

Second, the convergence of the bounds follows from the fact that every subsequence of $(l_f^r)_{r \in \mathbb{Z}_+}$ has a converging subsequence $(l_f^{r_k})_{k \in \mathbb{Z}_+}$ with limit $\pi(f)$, where $\pi \in \mathcal{P}$ (see Remark 16 and Theorem 3.5 in Ref. 38). By definition, $l_f \leq \pi(f)$; hence, taking limits in (81) shows that $l_f \geq \pi(f)$ and the result follows. The case of the upper bounds is identical. For details, see Corollary 3.6(i) in Ref. 38. \square

Remark 16 (LP computation). The bounds l_f^r and u_f^r in (80) are obtained by solving two linear programs (LPs) with $|S_r|$ variables, $|N_r|$ equality constraints, and $|S_r| + 3$ inequality constraints. LP solvers return the optimal value l_f^r (or u_f^r) and an optimal point $\pi^{*,r}$ such that $\langle f \rangle_{\pi^{*,r}} = l_f^r$ (or $\langle f \rangle_{\pi^{*,r}} = u_f^r$). The optimal points exist because the LPs are optimizations of a continuous function over a compact nonempty subset of $\mathbb{R}^{|S_r|}$.

Theorem 15 provides a general framework to obtain bounds (80) that can be used as approximations of stationary averages $\langle f \rangle_\pi$ with a quantifiable error given by (i)–(iii); furthermore, under the conditions in (iv), the approximations converge to $\langle f \rangle_\pi$ as the truncations approach the entire state space S of the reaction network.

1. The case of a unique distribution: Bounds and approximations

Throughout this section, we assume that the CME has a unique stationary solution π , i.e.,

$$\mathcal{P} = \{\pi\}.$$

In this case, the results of Theorem 15 can be strengthened. And in particular, the feasible points $\pi^r \in \mathcal{P}^r$ are good approximations of the

stationary solution in the sense that they converge to π in *weak**, as detailed in the following corollary.

Corollary 17 (Convergence of bounds and feasible points for a unique solution). Let us assume that the conditions of Theorem 15(iv) hold and, in addition, that $\mathcal{P} = \{\pi\}$ consists of a single stationary solution π . Then, the upper and lower bounds (80) converge to the average,

$$\lim_{r \rightarrow \infty} l_f^r = \lim_{r \rightarrow \infty} u_f^r = \langle f \rangle_\pi, \quad (85)$$

and any sequence of feasible points $(\pi^r)_{r \in \mathbb{Z}_+}$ belonging to the outer approximations $(\mathcal{P}^r)_{r \in \mathbb{Z}_+}$ (i.e., $\pi^r \in \mathcal{P}^r, \forall r \in \mathbb{Z}_+$) converges to π in *weak**,

$$\lim_{r \rightarrow \infty} \langle g \rangle_{\pi^r} = \langle g \rangle_\pi, \quad (86)$$

for any function g that satisfies (82).

Proof. This proof is similar to that of Theorem 15(iv). For full details, see Corollary 3.6(iii) in Ref. 38. \square

Remark 18. When w is normlike, convergence in *weak** implies convergence in total variation—see Appendix B of Ref. 38 for a proof.

Remark 19. Corollary 17 shows that, given a sufficiently large truncation, the feasible points π^r provide arbitrarily accurate approximations to π , yet with no quantifiable bound on the approximation error $\|\pi - \pi^r\|$.

a. *Bounding the stationary solution.* To obtain approximations with quantifiable errors, here we compute bounds on $\pi(x)$ for every state in the truncation by solving LPs (80) for various functions f . In particular, setting f to be the indicator function 1_x of a state x in S_r [cf. (3)] and noting that $\langle 1_x \rangle_{\pi_r} = \pi(x)$, we obtain the bounds

$$l_x^r \leq \pi(x) \leq u_x^r, \quad \forall x \in S_r, \quad (87)$$

where

$$l_x^r := l_{1_x}^r = \inf\{\pi^r(x) : \pi^r \in \mathcal{P}^r\}, \quad (88)$$

$$u_x^r := u_{1_x}^r = \sup\{\pi^r(x) : \pi^r \in \mathcal{P}^r\}. \quad (89)$$

As in (64) and (65), we then collect these bounds, pad them with zeros, and define two approximations for π ,

$$l^r := (l^r(x))_{x \in S}, \quad l^r(x) := \begin{cases} l_x^r & \text{if } x \in S_r \\ 0 & \text{if } x \notin S_r \end{cases}, \quad (90)$$

$$u^r := (u^r(x))_{x \in S}, \quad u^r(x) := \begin{cases} u_x^r & \text{if } x \in S_r \\ 0 & \text{if } x \notin S_r \end{cases}. \quad (91)$$

These approximations of π have controlled errors, as summarized in the following corollary.

Corollary 20 (Upper and lower bounding approximations of the unique solution). If Assumption 13 holds and $\mathcal{P} = \{\pi\}$, then the approximations l^r (90) and u^r (91) fulfill the following:

- (i) The approximations l^r and u^r bound π from below and above, respectively,

$$l^r(x) \leq \pi(x), \quad \forall x \in \mathcal{S}, \quad (92)$$

$$\pi(x) \leq u^r(x), \quad \forall x \in \mathcal{S}_r, \quad (93)$$

with quantified approximation errors ε_r^l and ε_r^u ,

$$\|l^r - \pi\| = 1 - l^r(\mathcal{S}_r) =: \varepsilon_r^l, \quad (94)$$

$$\begin{aligned} \|u^r - \pi\| &= \max\{u^r(\mathcal{S}_r) - 1 + m_r, m_r\} \\ &\leq \max\{u^r(\mathcal{S}_r) - 1 + c/r, c/r\} =: \varepsilon_r^u, \end{aligned} \quad (95)$$

where $\|\cdot\|$ denotes the total variation norm (66).

- (ii) As the truncation size r approaches infinity (and \mathcal{S}_r approaches \mathcal{S}), the approximation u^r converges pointwise to the unique π ,

$$\lim_{r \rightarrow \infty} u^r(x) = \pi(x), \quad \forall x \in \mathcal{S}, \quad (96)$$

and the approximation l^r converges to π in weak* (86).

Proof. The bounds (92) and (93) follow directly from Theorem 15. The error (94) follows from (92) and the fact that the total variation norm of an unsigned measure is its mass. Similarly, (95) follows from (76), (93), and

$$\begin{aligned} |u^r(A) - \pi(A)| &\leq \max\{u^r(A) - \pi|_r(A), \pi(A \cap \mathcal{S}_r^c)\} \\ &\leq \max\{u^r(\mathcal{S}_r) - 1 + m_r, m_r\} \\ &= \max\{|u^r(\mathcal{S}_r) - \pi(\mathcal{S}_r)|, |u^r(\mathcal{S}_r^c) - \pi(\mathcal{S}_r^c)|\}, \quad \forall A \subseteq \mathcal{S}. \end{aligned}$$

Theorem 15(iv) shows that l^r and u^r converge pointwise to π . For any f satisfying (82) and $r, r' \in \mathbb{Z}_+$, we have that

$$|\pi(f) - l^r(f)| \leq \sum_{x \in \mathcal{S}_r} |f(x)| (\pi(x) - l^r(x)) + 2 \sum_{x \notin \mathcal{S}_r} |f(x)| \pi(x).$$

Using the pointwise convergence of l^r and (84), we can pick r' such that the second sum is arbitrarily small and, subsequently, an r such that the first sum is arbitrarily small. Hence, the weak* convergence of l^r follows. See Theorem 4.1 in Ref. 38 for details. \square

Corollary 20 states that, for sufficiently large r , l^r and u^r are close to π . In contrast with the feasible points π^r , we can answer the question “is r sufficiently large?” by evaluating the errors ε_r^l (94) and ε_r^u (95). Since

$$\lim_{r \rightarrow \infty} \varepsilon_r^l = 0,$$

we will always find an approximation l^r that verifiably meets any given error tolerance by increasing r .

Remark 21. Although we have no proof that ε_r^u converges to zero (nor that u^r itself converges to π in total variation), all the examples we have encountered in practice exhibit convergence of u^r and ε_r^u .

To instead answer the question “when is r too small?” (i.e., to establish how large the partition \mathcal{S}_r must be to guarantee a given approximation error), the following proposition is of use.

Proposition 22 (Achievable approximation errors). Under the same conditions as in Corollary 20, the errors of the approximations l^r (90) and u^r (91) cannot be made smaller than the tail bound or the mass of the tail, respectively, i.e.,

$$\|l^r - \pi\| \geq \varepsilon_r, \quad \forall r \in \mathbb{Z}_+, \quad (97)$$

$$\|u^r - \pi\| \geq m_r, \quad \forall r \in \mathbb{Z}_+, \quad (98)$$

where $m_r = \pi(\mathcal{S}_r^c) \leq \varepsilon_r = c/r$.

Proof. The inequality (98) follows directly from (95). For (97), recall that there exists at least one optimal point $\pi^{*,r}$ such that $\pi^{*,r}(x) = l^r(x)$ for any given $x \in \mathcal{S}_r$ (Remark 16). It is straightforward to verify that

$$\frac{1 - \varepsilon_r}{\pi^{*,r}(\mathcal{S}_r)} \pi^{*,r} \in \mathcal{P}^r,$$

which implies $\pi^{*,r}(\mathcal{S}_r) = 1 - \varepsilon_r$ [due to the minimality of $\pi^{*,r}(x)$]. Since l^r bounds from below all feasible points of \mathcal{P}^r , it follows that $l^r(\mathcal{S}_r) \leq \pi^{*,r}(\mathcal{S}_r) = 1 - \varepsilon_r$. Combined with (94), this gives (97). \square

In other words, the approximation error of the lower bounds is no smaller than the tail bound, whereas that of the upper bounds is no smaller than the tail mass.

Remark 23. The inequalities (97) and (98) are sharp for birth-death processes with $w(x) = x^\alpha$ [see (67) and (68)].

b. Approximating marginal distributions. For high-dimensional state spaces, we are often interested in marginal distributions rather than the full multivariate solution π defined on \mathcal{S} . A marginalization is associated with a partition of the state space into a collection of disjoint subsets,

$$\{A_i\}_{i \in \mathcal{I}}, \quad \cup_{i \in \mathcal{I}} A_i = \mathcal{S}, \quad A_i \cap A_j = \emptyset, \quad \forall i \neq j \in \mathcal{I},$$

and the marginal distribution is defined with respect to each subset,

$$\hat{\pi}(i) = \pi(A_i) = \sum_{x \in A_i} \pi(x), \quad \forall i \in \mathcal{I}. \quad (99)$$

Because $\{A_i\}_{i \in \mathcal{I}}$ is a partition of \mathcal{S} , $\hat{\pi}$ is a probability distribution on \mathcal{I} .

Typically, we are interested in marginalizing the distribution of a reaction network with n species (and state space $\mathcal{S} = \mathbb{N}^n$) over a subset of species. For instance, if we are interested in the molecule counts of species k , we consider the following (infinite) set of subsets:

$$\{A_i\}_{i \in \mathbb{N}}, \quad \text{where } A_i := \mathbb{N}^{k-1} \times \{i\} \times \mathbb{N}^{n-k}, \quad (100)$$

whose union trivially recovers the entire state space. Associated with this set $\{A_i\}_{i \in \mathbb{N}}$, we then have the marginal distribution $\hat{\pi}$

$$\hat{\pi}(i) = \pi(\{x \in \mathbb{N}^n : x_k = i\}), \quad \forall i \in \mathbb{N},$$

which, in this case, corresponds to the (univariate) distribution describing the molecule counts of the k th species.

The marginal distribution $\hat{\pi}$ can also be bounded and approximated following a similar procedure to the one described above

for the full distribution. Using the indicator functions 1_{A_i} as the functions f , we solve the analogous LPs,

$$\hat{l}_i^r = \inf\{\pi^r(A_i) : \pi^r \in \mathcal{P}^r\}, \quad \forall i \in \mathcal{I}_r, \quad (101)$$

$$\hat{u}_i^r = \sup\{\pi^r(A_i) : \pi^r \in \mathcal{P}^r\}, \quad \forall i \in \mathcal{I}_r, \quad (102)$$

for all the subsets A_i that intersect with the truncation, i.e., $\mathcal{I}_r = \{i \in \mathcal{I} : A_i \cap \mathcal{S}_r \neq \emptyset\}$.

As before, we construct two approximations by padding (101) and (102) with zeros,

$$\hat{l}^r := (\hat{l}^r(i))_{i \in \mathcal{I}}, \quad \hat{l}^r(i) := \begin{cases} \hat{l}_i^r & \text{if } i \in \mathcal{I}_r \\ 0 & \text{if } i \notin \mathcal{I}_r \end{cases}, \quad (103)$$

$$\hat{u}^r := (\hat{u}^r(i))_{i \in \mathcal{I}}, \quad \hat{u}^r(i) := \begin{cases} \hat{u}_i^r & \text{if } i \in \mathcal{I}_r \\ 0 & \text{if } i \notin \mathcal{I}_r \end{cases}, \quad (104)$$

which are the analogs for the marginal distribution of the approximations to the entire distribution (90) and (91) and have similar (but not identical) properties, as summarized in the following two corollaries.

Corollary 24 (Lower bounding approximation of the marginal distribution). *Let us assume that the conditions of Corollary 20 hold. If $\{A_i\}_{i \in \mathcal{I}}$ is a partition of \mathcal{S} , then the associated marginal distribution $\hat{\pi}$ (99) is lower bounded by the approximation \hat{l}^r defined in (103),*

$$\hat{l}^r(i) \leq \hat{\pi}(i), \quad \forall i \in \mathcal{I},$$

with approximation error

$$\|\hat{l}^r - \hat{\pi}\| = 1 - \hat{l}^r(\mathcal{I}_r) =: \hat{\varepsilon}_r^l. \quad (105)$$

Furthermore, as the truncation size is increased ($r \rightarrow \infty$ and \mathcal{S}_r approaches \mathcal{S}), \hat{l}^r converges to $\hat{\pi}$ in total variation.

Proof. This proof is analogous to that of Corollary 20 except that one needs to use Corollary 4.3 in Ref. 38 [with $g(x, y) = 1_{A_i}(x)$ instead of Theorem 4.1 in Ref. 38]. \square

Corollary 25 (Convergent approximation of the marginal distribution). *Under the same conditions as in Corollary 24, \hat{u}^r defined in (104) approximates the marginal distribution $\hat{\pi}$ with error bounded by*

$$\|\hat{u}^r - \hat{\pi}\| \leq \max\{\hat{u}^r(\mathcal{I}_r) - 1 + m_r, m_r\} \leq \max\{\hat{u}^r(\mathcal{I}_r) - 1 + c/r, c/r\} =: \hat{\varepsilon}_r^u. \quad (106)$$

Furthermore, as the truncation \mathcal{S}_r approaches \mathcal{S} , \hat{u}^r converges pointwise to $\hat{\pi}$,

$$\lim_{r \rightarrow \infty} \hat{u}^r(i) = \hat{\pi}(i), \quad \forall i \in \mathcal{I}.$$

Proof. The proof is analogous to that of Corollary 20. \square

Note that \hat{u}^r does provide a controlled approximation of the marginal distribution as it is a pointwise convergent approximation to $\hat{\pi}$ with a guaranteed, computable error bound $\hat{\varepsilon}_r^u$ (106).

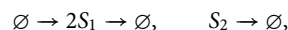
Remark 26 (Upper bounds for the marginal distribution). *The approximation $\hat{u}^r(i)$ bounds the marginal $\hat{\pi}(i)$ if and only if $A_i \cap \mathcal{S}_r^c = \emptyset$, i.e., when the set A_i is fully contained inside the truncation \mathcal{S}_r . Hence, $\hat{u}^r(i)$ does not provide an upper bound if the truncation does not include all marginalized variables. However, using the fact that the probability mass of $A_i \cap \mathcal{S}_r^c$ is bounded by the mass of the tail m_r (76), we have the following easy (but loose) upper bounds,*

$$\hat{\pi}(i) \leq \hat{u}^r(i) + c/r, \quad \forall i \in \mathcal{I}_r. \quad (107)$$

2. Nonuniqueness, ergodic distributions, and a uniqueness test

Theorem 15 shows that our LP optimization over the polytopes \mathcal{P}^r yields bounds on the stationary averages, even if there are multiple stationary solutions. In the nonunique case, however, the gap between the lower bounds and the upper bounds will reflect the fact that the extreme points of $\pi \mapsto \langle f \rangle_\pi$ over \mathcal{P} can be achieved by different solutions in the polytope. Yet it is possible to characterize further the set of solutions and the extreme points in terms of the ergodic distributions of the CME and use this description to turn our LP approach into a test of uniqueness, as we show below.

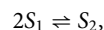
To see how multiple stationary solutions of the CME can arise, consider the simple reaction network



with mass action kinetics. It is clear that its state space $\mathcal{S} = \mathbb{N}^2$ decomposes into three disjoint sets,

$$\mathcal{S} = \{(x_1, 0) : x_1 \in \mathbb{N} \text{ is odd}\} \cup \{(x_1, 0) : x_1 \in \mathbb{N} \text{ is even}\} \cup \{(x_1, x_2) : x_1 \in \mathbb{N}, x_2 \in \mathbb{Z}_+\} =: \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{T},$$

where \mathcal{C}_1 and \mathcal{C}_2 are closed communicating classes and \mathcal{T} contains the remaining states. A set $\mathcal{C} \subseteq \mathcal{S}$ is a closed communicating class³⁹ if the chain can transit between any pair of states in \mathcal{C} but cannot leave \mathcal{C} . Another common source of multiple closed communicating classes is conservation laws in reaction networks.⁵⁴ For instance, the reactions



conserve the quantity $n = x_1 + 2x_2$. Hence, there exists a different closed communicating class for every $n \in \mathbb{N}$.

The closed communicating classes are intimately related to the stationary solutions, as summarized in the following theorem that compiles some facts that are broadly known in the literature.

Theorem 27 (Ergodic distributions and communicating classes¹⁵). *Consider a reaction network (1) with rate matrix Q satisfying (2)–(5), assume Q is regular, and decompose the state space as*

$$\mathcal{S} = (\cup_j \mathcal{C}_j) \cup \mathcal{T}, \quad (108)$$

where \mathcal{C}_j are closed communicating classes and \mathcal{T} contains the remaining states.

- (i) For each C_j , there is at most one stationary solution π_j ; hence, $\pi_j(C_j) = 1$. Whenever it exists, π_j is known as the ergodic distribution associated with C_j .
- (ii) Let \mathcal{J} be the set of indexes j of the ergodic distributions π_j . The set of stationary solutions \mathcal{P} (9) is the set of convex combinations of the ergodic distributions,

$$\mathcal{P} = \left\{ \sum_{j \in \mathcal{J}} \theta_j \pi_j : \theta_j \geq 0 \quad \forall j \in \mathcal{J}, \quad \sum_{j \in \mathcal{J}} \theta_j = 1 \right\}. \quad (109)$$

Proof. If Q is regular, the stationary solutions of the CME are the stationary distributions of the chain (Theorem 30 in Appendix A). Using this fact, (i) can be found in most books on continuous-time chains [e.g., Ref. 39 (Theorem 3.5.2)], and (ii) is given in Ref. 15 (Theorem 3.4). \square

Theorem 27 states that the ergodic distributions π_j are orthogonal to each other and that they are the extreme points of the convex polytope \mathcal{P} (9) of stationary solutions of the CME. Because \mathcal{P} is contained in the non-negative orthant of ℓ^1 , it follows that each face of the non-negative orthant contains at most one of the ergodic distributions. Using this fact, we obtain a computational test of the uniqueness of stationary solutions, as summarized in the following corollary.

Corollary 28 (A uniqueness test). *If Assumption 13 holds and Q is regular, then*

$$\mathcal{P} = \{\pi\} \iff \exists x \in \mathcal{S}, \quad r \geq 1 : l'(x) > 0.$$

Proof. The proof is by contradiction. Suppose that \mathcal{P} is not a singleton. Theorem 27(ii) implies that \mathcal{P} contains two or more ergodic distributions, π_j , each associated with a different closed communicating class, C_j . Let us consider a state $x \in C_1$. Since the classes C_j are disjoint, then $\pi_j(x) = 0, \forall j \neq 1$, which contradicts the lower bound property of l' [Theorem 15(i)]. Hence, \mathcal{P} must be a singleton. The converse follows from the convergence of the bounds in Corollary 20 (ii). \square

In other words, if π is unique, then the lower bound of any (and all) states in the support of π is nonzero for sufficiently large r . Conversely, finding a single nonzero lower bound for any $x \in \mathcal{S}$ provides a proof of uniqueness of the distribution. Hence, if there is more than one ergodic distribution, all the lower bounds are zero for all states in the state space \mathcal{S} .

When π is unique, l' or u^r are good approximations of the stationary solution. However, this is not so in the nonunique case. Indeed, it is easy to show that, in the nonunique case, the lower and upper bounds are always loose: Corollary 28 shows that the lower bounds are trivially zero everywhere ($l' = 0$), whereas Theorems 15(ii) and 27 imply that for large r , the mass of u^r will be no smaller than the number of ergodic distributions,

$$\liminf_{r \rightarrow \infty} u^r(\mathcal{S}) \geq |\mathcal{J}|; \quad (110)$$

hence, the upper bound is not tight.

However, our LP framework can still be used to obtain approximations of the ergodic distributions by using sequences of feasible

points π^r . To this end, we require the following generalization of Corollary 17.

Corollary 29 (Convergent approximations of ergodic distributions). *Let f be any function that satisfies (82) (i.e., the growth of f is strictly dominated by that of the normlike function w), and let $(\pi^{*,r})_{r \in \mathbb{Z}_+}$ be a sequence of optimal points such that*

$$\pi^{*,r} \in \mathcal{P}^r \quad \text{and} \quad \langle f \rangle_{\pi^{*,r}} = u_f^r = \sup\{\langle f \rangle_{\pi} : \pi \in \mathcal{P}^r\},$$

for all $r \in \mathbb{Z}_+$.

- (i) If there exists a unique point $\pi^* \in \mathcal{P}$ such that

$$\langle f \rangle_{\pi^*} = \sup\{\langle f \rangle_{\pi} : \pi \in \mathcal{P}\}, \quad (111)$$

then the sequence of optimal points $\pi^{*,r}$ converges to π^* in weak* as $r \rightarrow \infty$.

- (ii) If f is the indicator function of a state (or of a subset) that is contained in a closed communicating class C_j with associated ergodic distribution π_j , then the sequence of optimal points $(\pi^{*,r})_{r \in \mathbb{Z}_+}$ converges to π_j in weak* as $r \rightarrow \infty$.

Proof. The proof of (i) is similar to that of Theorem 15(iv). See Corollary 3.6(ii) and Remark 3.7 in Ref. 38 for details. (ii) follows from (i), Theorem 27(ii), and the fact that indicator functions satisfy (82). \square

Corollary 29 provides a rationale for how to use our computational framework to obtain approximations of the ergodic distributions π_j in the nonunique case. Importantly, we do not need to know a priori what the closed communicating classes are. Using the indicator function for a chosen state x , we obtain the sequence of optimal points $\pi^{*,r}$ satisfying $\pi^{*,r}(x) = u^r(x)$. Should x belong to a closed communicating class C_j with ergodic distribution π_j , Corollary 29(ii) shows that $\pi^{*,r}$ will converge to π_j as r tends to infinity. Indeed, by looking at the states for which $\pi^{*,r}(x) > 0$, we can in principle deduce which communicating class C_j (if any) the state belongs to. Once the class is known, we replace \mathcal{S} with C_j and proceed as for the unique case to obtain bounds on π_j . See Sec. V C for an example of the application of this procedure.

3. Computational implementation and numerical considerations

Let us consider a given reaction network with rational propensities. In order to obtain approximations of its stationary solutions with controlled error smaller than a tolerance ϵ , we proceed as follows:

1. Verify the existence of stationary solutions π and the finiteness of their moments (Assumption 1) using a Foster-Lyapunov criterion (Theorem 31 in Appendix B).
2. Choose a normlike rational function w and define the truncations \mathcal{S}_r as the sublevel sets (74) controlled by r .

We have found it best to choose functions w that define truncations that cover most of the probability mass and that tend quickly to infinity so that the size of the truncation grows slowly with r . For example, if we take $w(x) = x^\alpha$, then higher values of α induce smaller truncation sizes $|\mathcal{S}_r| \approx \sqrt[r]{r}$. To guide the selection of w , one can run the scheme with various w to gain information about the shape of the distribution.

- Use the SDP approach of Sec. III B to find a moment bound (75) with bounding constant c satisfied by all stationary solutions (Assumption 13). In particular, we employ YALMIP,⁴⁶ SDPA-GMP,⁴⁷ and mpYALMIP⁴⁸ to formulate and solve the SDP (48) with $f := w$. See details in Sec. III B 2.
- Choose an initial r based on the achievable errors established in Proposition 22. To guarantee an error smaller than our tolerance ϵ , we must choose an initial $r > c/\epsilon$.
- Solve the LPs (88) to obtain the upper and lower approximations l_r (90) and u_r (91). Here, we use the dual simplex algorithm of CPLEX V12.6.3⁵² to solve the LPs. The tool YALMIP⁴⁶ is convenient to formulate the LPs, but if speed is a priority, the model can be fed straight to the solver to avoid computational overheads.
For large truncations, the coefficients in the constraints of the LPs span many orders of magnitude, leading to round-off errors in double-precision arithmetic and poor solver performance. One way to ameliorate this issue is to scale the decision variables; in particular, scaling $\pi^r(x)$ by $-q(x, x)$ or $w(x)$ often significantly improves solver performance.
- Evaluate the error of the approximations l_r and u_r using (94) and (95), respectively. If the error is larger than our tolerance ϵ , we increase the truncation size r and return to the previous step.
- In addition to approximating the full distribution, we can apply the above steps to compute other measures of interest by changing the LPs and associated errors in steps 5 and 6:
 - If we want to approximate a marginal distribution, we solve the LPs (101) and (102) and quantify the error using (105) and (106).
 - If we are interested in a particular stationary average $\langle f \rangle_\pi$, we instead solve the LPs (80) and control the error using the bounds in Theorem 15(i)–(iii).
- As the particular reaction network could have several stationary solutions, we check in step 5 for nontrivial lower bounds, $l_r(x) > 0$. If we find one such bound, the solution is unique (Corollary 28). Otherwise, we investigate further the uniqueness question by increasing r and recomputing the lower bounds to examine the presence of communicating classes as discussed in Corollary 29.

Our computations were carried out on a desktop computer with a 3.5 GHz processor and 16 GB of RAM.

V. APPLICATION TO BIOLOGICAL EXAMPLES

We now present the application of the methodology to three examples. First, we showcase how to obtain tight bounds on the stationary solution (and marginals) of a two-dimensional toggle switch. Second, we consider a model of bursty gene expression with negative feedback, through which we explore the capabilities of our method to deal with promoter switching noise. Third, we demonstrate the application of our methods to the nonunique case with a dimerization network. The code used to compute the approximations and bounds for this last example is available at Ref. 55.

A. A toggle switch

Toggle switches are common motifs in many cell-fate decision genetic circuits.^{16,56,57} A simple such circuit consists of two mutually repressing genes.⁵⁶ In particular, we consider the asymmetric case with mutual repression modeled via Hill functions and dilution/degradation modeled via linear decay,



The state space of the CME is $x \in \mathcal{S} = \mathbb{N}^2$ with $x = (x_1, x_2)$, where x_1 and x_2 denote the number of protein P_1 and P_2 , respectively, and the propensities of the reactions are

$$\begin{aligned} a_1(x) &= \frac{k_1}{1 + (x_2/\theta)^3}, & a_2(x) &= k_2 x_1, \\ a_3(x) &= \frac{k_3}{1 + x_1}, & a_4(x) &= k_4 x_2, \end{aligned} \quad (113)$$

where $k_i > 0$ are kinetic constants and $\theta > 0$ is the dissociation constant of P_2 .

We follow the steps detailed in Sec. IV B 3 to obtain bounds and approximations for this reaction network. First, we show that a stationary solution π exists and that all of the moments of every solution are finite using a Foster-Lyapunov criterion (Appendix B).

We pick the normlike function

$$w(x) := (x_1 + x_2)^6, \quad \forall x \in \mathbb{N}^2,$$

and compute the moment bound

$$\langle w \rangle_\pi \leq c := 4.48 \times 10^8$$

by solving the SDP (48) with $d := 10$ and $f := w$ (solver time = 3.6 min).

We then solve the LPs (88) and (89) and compute the bounding approximations u^r and l^r of the stationary solutions. The fact that the lower bounds l^r are nonzero provides us with a proof of uniqueness of the stationary solution. Figure 4 shows the bounds for small ($r = 45$) and large ($r = 75$) state space truncations. The maximum absolute discrepancies are found near the modes, and by increasing the size of the truncation, the upper and lower bounds become nearly indistinguishable—the maximum discrepancy drops under 10^{-4} [Fig. 4(d)]. Overall, the total approximation error including the tail is less than 2.6×10^{-3} , as given by (94) and (95).

We have also used our method to obtain approximations on the marginal distributions of the number of proteins P_1 (x_1) and P_2 (x_2) (Sec. IV B 1). The results in Fig. 5(a) show that the bounds get tighter for truncations of increasing r [although, as discussed in Remark 26, the upper bound (107) remains loose]. Note, however, that the approximation \hat{u}^r in Fig. 5(b) rapidly approaches the Gillespie numerical simulations. Figure 5(c) shows that the errors of both \hat{l}^r and \hat{u}^r can be made arbitrarily small by increasing the truncation size (Corollaries 24 and 25).

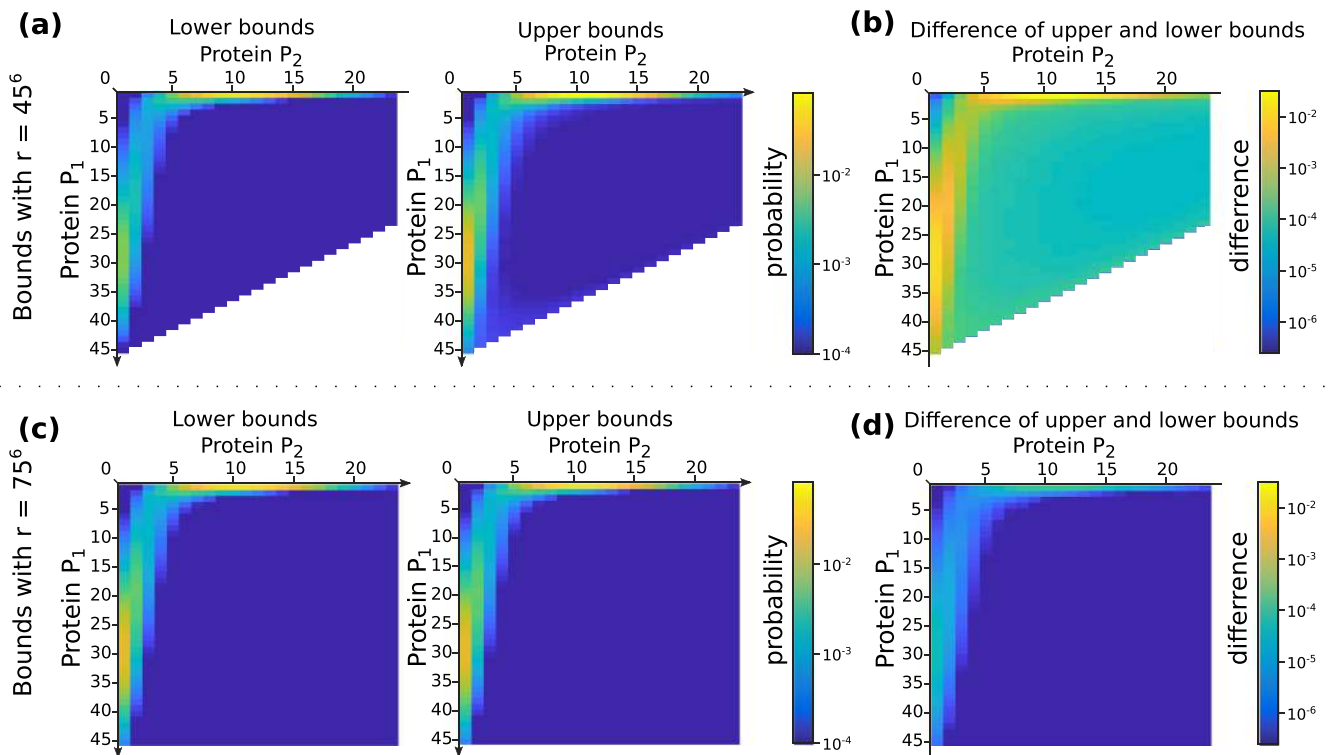


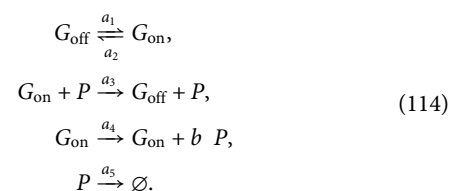
FIG. 4. Bounds on the stationary solution of the toggle switch (112) as the state space truncation is increased. (a) Lower and upper bounds l and u for the truncation $S_{46^6} = \{(x_1 + x_2)^6 < r = 46^6\}$ (990 equations involving 1035 states). The white areas indicate states outside of the truncation. Approximation errors of the lower and upper bounding approximations: $\epsilon_r^l = 0.31$ and $\epsilon_r^u = 0.41$, respectively. In total, 2070 bounds were computed (solver time = 5 min, average of 0.15 s per bound). (b) Gap between upper and lower bounds: the largest uncertainties occur near the modes. (c) Same as (a) but with truncation parameter increased to $r = 75^6$ (2775 equations involving 2850 states). The upper and lower bounds are visually indistinguishable, with approximation errors $\epsilon_r^l = 2.5 \times 10^{-3}$ and $\epsilon_r^u = 2.6 \times 10^{-3}$. In total, 5700 bounds were computed (solver time = 64 min, average of 0.7 s per bound). (d) The maximum absolute gap between bounds is less than 10^{-4} . Parameters: $\theta = 1$, $k_1 = 30$, $k_2 = k_4 = 1$, and $k_3 = 10$.

Finally, we apply the method to chart the change in the stationary solution as a function of a parameter. In particular, the dissociation constant of protein P_2 (θ) can be thought of as a bifurcation parameter: increasing θ allows for higher expression of protein P_1 . Figure 6 presents the lower bounds \hat{l}_r on the marginals of both proteins. At small values of θ , we observe a single population with high numbers of P_2 repressing P_1 . At large values of θ , the opposite happens: the population we observe has high numbers of P_1 repressing P_2 . For intermediate θ , we observe coexistence of both populations. Indeed, we find that the modes of the marginal distributions are in good correspondence with the stable solutions of the deterministic steady-state rate equations.

B. Bursty gene expression with negative feedback

As a second example, consider a model of bursty production of a protein that regulates (negatively) its own expression. The model⁵⁸ involves a promoter that switches between active (G_{on}) and inactive (G_{off}) states, and the protein P it encodes. When the promoter is on, the protein is expressed in bursts of size b , a geometrically distributed random variable⁵⁹ with mean $\langle b \rangle$. The protein represses its

own production by switching off the promoter,



The state space of the CME is $x = (x_1, x_2) \in \mathcal{S} = \{0, 1\} \times \mathbb{N}$, where $x_1 = \{0, 1\}$ is a binary variable describing the off/on state of the promoter and $x_2 \in \mathbb{N}$ represents the protein count. The propensities are

$$\begin{aligned} a_1(x) &= k_1(1 - x_1), \quad a_2(x) = k_2x_1, \\ a_3(x) &= k_3x_2x_1, \quad a_4(x) = k_4x_1, \quad a_5(x) = k_5x_2, \end{aligned}$$

where the $k_i > 0$ are reaction rate constants. In Appendix B, we show that the network has a unique stationary solution π and that all of its moments are finite.

This example provides an interesting test case for SDP methods since the protein noise is particularly large: $CV(x_2)$, the coefficient of

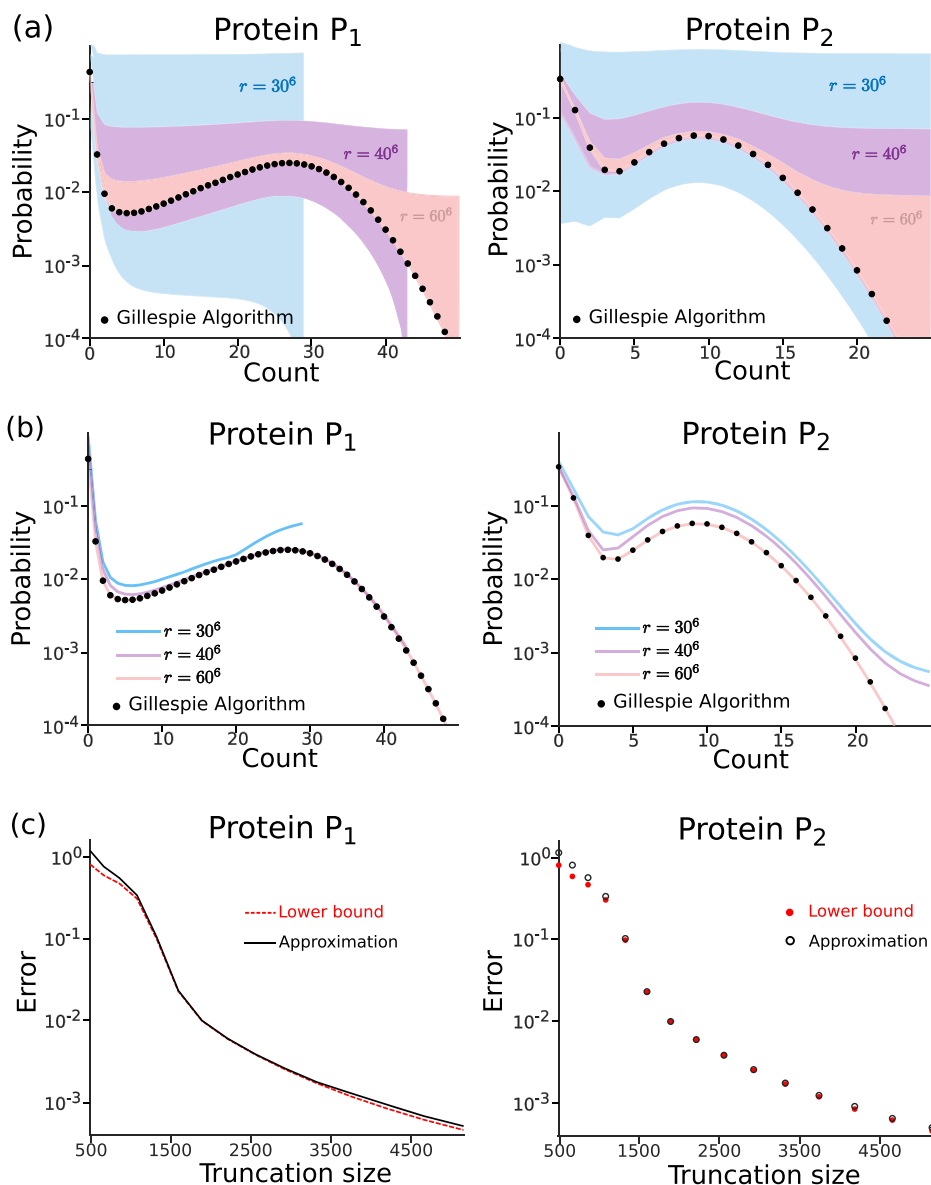


FIG. 5. Marginal distributions of the toggle switch (112). (a) Tight lower bounds \hat{l}^r and loose upper bounds (107) on the marginal distributions of both proteins computed for increasing state space truncations $r = 30^6$ (cyan), 40^6 (purple), and 60^6 (pink). Overall, 520 bounds were computed (solver time = 86.6 s, average of 0.17 s per bound). For comparison, we show simulations performed using the Gillespie algorithm with 10^8 samples (black dots). (b) Controlled approximations \hat{u}^r for increasing state space truncations $r = 30^6$ (cyan), 40^6 (purple), and 60^6 (pink). (c) Approximation error $\hat{\epsilon}_r^l$ of the lower bound \hat{l}^r (red) and bound $\hat{\epsilon}_r^u$ on error of \hat{u}^r (black). In total, 3900 bounds were computed (solver time = 48.8 min, average of 0.75 s per bound). Parameters as in Fig. 4.

variation of x_2 , grows⁵⁸ with the burst size $\langle b \rangle$. Therefore, we expect that getting tight bounds for the CV will entail the use of a large number of moment equations. To investigate the effect of such large noise on the efficacy of our SDP method, we compute the following bounds:

$$\frac{\sqrt{L_{x_2}^d - (U_{x_2}^d)^2}}{U_{x_2}^d} \leq \text{CV}(x_2) \leq \frac{\sqrt{U_{x_2}^d - (L_{x_2}^d)^2}}{L_{x_2}^d}, \quad (115)$$

where we use (47) and (48) and we append the following equalities to our SDP (45),

$$x_1 \in \{0, 1\} \implies \langle x_1^{\alpha_1} x_2^{\alpha_2} \rangle_\pi = \langle x_1 x_2^{\alpha_2} \rangle_\pi \alpha_1 > 0, \alpha_2 \geq 0.$$

Figure 7(a) shows how the bounds (115) get tighter as we increase the number of moment equations in our SDP calculations. As expected, for small mean burst sizes ($\langle b \rangle = 1$), the bounds become tight with 10 moment equations, but tightening the bounds becomes difficult when the burst size is larger ($\langle b \rangle = 10, 100$).

Computing tight bounds for large $\langle b \rangle$ with the naive SDP approach would thus require a prohibitive number of moment equations. However, we can apply the LP method of Sec. IV to overcome this limitation. To do this, use SDP to compute a (cheaper) loose upper bound on the sixth moment $L_{x_2}^{13} = 2.3 \times 10^{13} \leq \langle x_2^6 \rangle_\pi \leq U_{x_2}^{13} = 4.5 \times 10^{13}$ (time = 1 min per bound). We then set $w(x) := x_2^6$ and $c := U_{x_2}^{13}$ in Theorem 15 to obtain

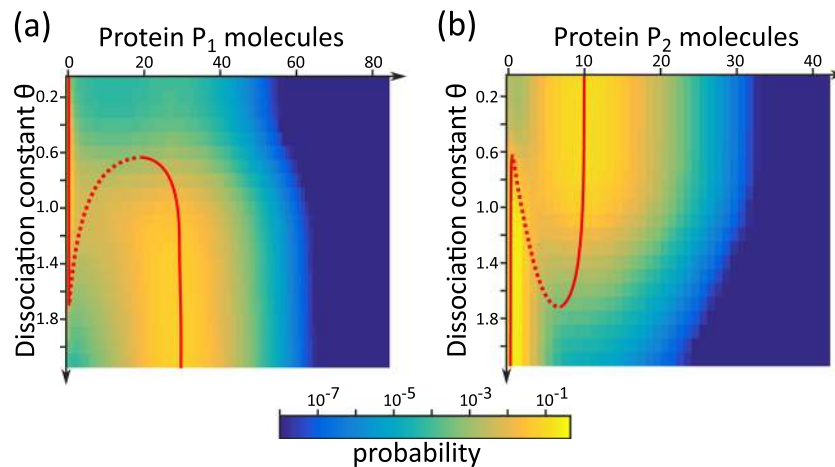


FIG. 6. Toggling the switch: deterministic vs stochastic. As the promoter dissociation constant θ is increased, the system switches from a state where overexpression of P_2 represses P_1 (low θ) to the reverse state where high P_1 represses P_2 (high θ). In the deterministic case, the two states (stable fixed points, solid red lines) coexist at intermediate θ and can be reached from different initial conditions separated by a third unstable steady state (dotted red line). For the stochastic model (112), we compute the marginal stationary probabilities (heatmap) $\hat{\pi}(x_1; \theta)$ in (a) and $\hat{\pi}(x_2; \theta)$ in (b) for different θ values and observe good correspondence of the modes of the distributions with the deterministic steady states. Each marginal is approximated by lower bounds computed using $r = 83^6$ (3403 equations involving 3486 states). In total, 3486 bounds were computed to obtain the full bifurcation diagram (θ increased in steps of 0.15): solver time = 62 min, average of 1 s per bound. Approximation error: $\hat{\epsilon}_i^l \leq 3 \times 10^{-3}, \forall \theta$. All parameters (other than θ) as in Fig. 4.

$$\begin{aligned} \hat{l}_{x_2}^r &\leq \langle x_2 \rangle_{\pi} \leq u_{x_2}^r + \frac{c}{r^{5/6}}, \\ \hat{l}_{x_2}^r &\leq \langle x_2^2 \rangle_{\pi} \leq u_{x_2}^r + \frac{c}{r^{2/3}}, \end{aligned}$$

which we combine as in (115) to obtain much tighter bounds on $CV(x_2)$. Figure 7(b) shows the convergence of these tight bounds for $CV(x_2)$ with $\langle b \rangle = 100$ as the truncation size is increased. These results exemplify the fact that it is enough to obtain loose SDP bounds on a higher order moment in order to obtain arbitrarily tight LP bounds on lower order moments (Theorem 15).

Finally, we exemplify in Fig. 7(c) another use of our capability to bound marginal distributions following the steps in Sec. IV B 3. In this case, we marginalize over the on/off promoter variable (x_1) and we compute upper bounds \hat{u}^r and lower bounds \hat{l}^r [visually indistinguishable in Fig. 7(c)] on $\hat{\pi}(x_2)$, the distribution of protein counts, for three different burst sizes, $\langle b \rangle = 1, 10, 100$. As expected, the protein distribution widens considerably (yet still with light tails) as the burst size increases. To compute these bounds, we set

$$A_i = \{0, 1\} \times \{i\}, \quad \forall i \in \{i \in \mathbb{N} : i^6 < r\} =: \mathcal{I}_r,$$

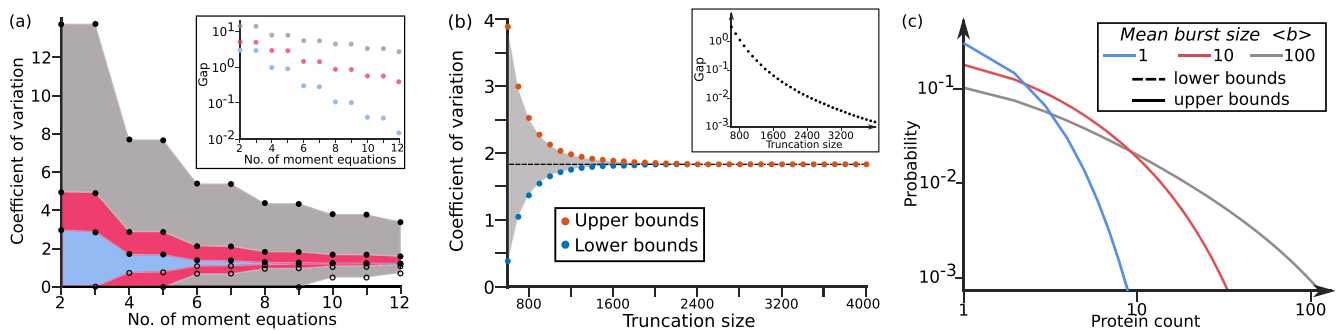


FIG. 7. Bounds for a bursty gene expression model with negative feedback (114). (a) Lower bounds (open circles) and upper bounds (filled circles) on the CV computed via SDP as the order of the approximation d is increased (no. of moment equations = $d - 1$) for different burst sizes, $\langle b \rangle = 1$ (cyan), 10 (crimson), and 100 (gray). Color shadings indicate the gap between bounds for different $\langle b \rangle$. In total, 44 bounds were computed for each $\langle b \rangle$ (solver time = 10 min, average of 13 s per bound). Inset: the gap decreases with increasing number of moment equations, albeit more slowly for larger $\langle b \rangle$. (b) For large mean burst size ($\langle b \rangle = 100$), the upper and lower bounds on the CV can be tightened using LPs with state-space truncations of increasing size (gap shaded in gray). In total, 140 bounds were computed (solver time = 11 min, average of 5 s per bound). Inset: the gap can be made arbitrarily small by increasing the truncation size. (c) Lower bounds \hat{l}^r and upper bounds \hat{u}^r (visually indistinguishable) on the marginal distribution of the protein for different burst sizes (with $r = 10^6, 55^6, 500^6$ for $\langle b \rangle = 1, 10, 100$, respectively). The total variation error is smaller than 4×10^{-3} in all cases. Overall, 1130 bounds were computed for all burst sizes (solver time = 7 min, average of 0.4 s per bound). Parameter values as in Ref. 58: $k_3 = k_4 = 10, k_1 = k_2 = k_5 = 1$.

and solve the LPs (101) and (102) to obtain (103) and (104). Note that this marginalization is over the *complete*, untruncated domain of the marginalized variable $x_1 = \{0, 1\}$. As a result, \hat{u}^r do provide upper bounds in this case (Remark 26).

C. Dimerization network with multiple stationary distributions

To illustrate the use of our method on CMEs with multiple stationary solutions (see Sec. IV B 2), we consider the reversible dimerization network



with state space $\mathcal{S} = \mathbb{N}$ and mass action kinetics: $a_1(x) := k_1$ and $a_2(x) := k_2 x_1(x_1 - 1)$. The reactions preserve the parity of the number of molecules; hence, the even numbers \mathcal{C}_0 and the odd numbers \mathcal{C}_1 are closed communicating classes. Furthermore, because the network obeys detailed balance,²³ it is straightforward to obtain analytical expressions for the two ergodic distributions,

$$\pi_i(x) = \frac{1}{Z_i} \frac{\mu^x}{x!}, \quad \forall x \in \mathcal{C}_i, \quad i = 0, 1, \quad (117)$$

where $\mu := \sqrt{k_1/k_2}$ and the normalizing constants are given by $Z_0 := \cosh(\mu)$ and $Z_1 := \sinh(\mu)$.

To illustrate the use of our tools in such a nonunique case, suppose we were not aware of the above facts and instead apply our computational procedure. First, we use a Foster-Lyapunov criterion and find that the rate matrix is regular; at least one stationary solution exists; all of the moments of each stationary solution are finite; and, for each closed communicating class, there exists an ergodic distribution with support in that class (Appendix B).

Next, we obtain bounds on the stationary solutions. Using the normlike function

$$w(x) := x^7, \quad \forall x \in \mathbb{N},$$

we solve the SDP (48) with $d := 7$ and $f := w$ (solver time = 3 s) to obtain the moment bound,

$$\langle w \rangle_\pi \leq c = 4.8814 \times 10^7.$$

We then solve the LPs (88) and (89) to compute upper and lower bounds, u^r and l^r [Fig. 8(a)]. Note that the lower bounds remain trapped at zero, indicating that the stationary solution is nonunique (as given by Theorem 28). Furthermore, we observe that the mass of the upper bounds approaches two as the truncation grows [i.e., $u^r(\mathcal{S}_r) \rightarrow 2$ as $r \rightarrow \infty$], hinting that there exist *two* ergodic distributions corresponding to two closed communicating classes [c.f. (110)].

To identify the communicating classes, we note that the optimal points $\pi_0^{*,r}$ and $\pi_1^{*,r}$ of the linear programs,

$$\sup\{\pi^r(0) : \pi^r \in \mathcal{P}^r\}, \quad \sup\{\pi^r(1) : \pi^r \in \mathcal{P}^r\}, \quad (118)$$

approach distributions with support on the even and odd numbers, respectively [Fig. 8(b)]. These observations indicate that the communicating classes are the even numbers \mathcal{C}_0 and the odd numbers \mathcal{C}_1 (Corollary 29).

To verify this claim, we compute two different sets of upper and lower bounds over \mathcal{C}_0 and \mathcal{C}_1 separately. When computed over the subsets \mathcal{C}_0 and \mathcal{C}_1 separately, the lower bounds are nonzero, thus attesting to the uniqueness of each stationary distribution over its communicating class. Using $r = 25^7$, the total variation error of the lower bounds (94) is approximately 8×10^{-3} , while that of the upper bounds (95) is bounded above by 8×10^{-3} . In fact, the actual errors of the upper bounds computed from the analytical expressions (117) are $< 10^{-4}$, i.e., substantially smaller than the guaranteed bound.

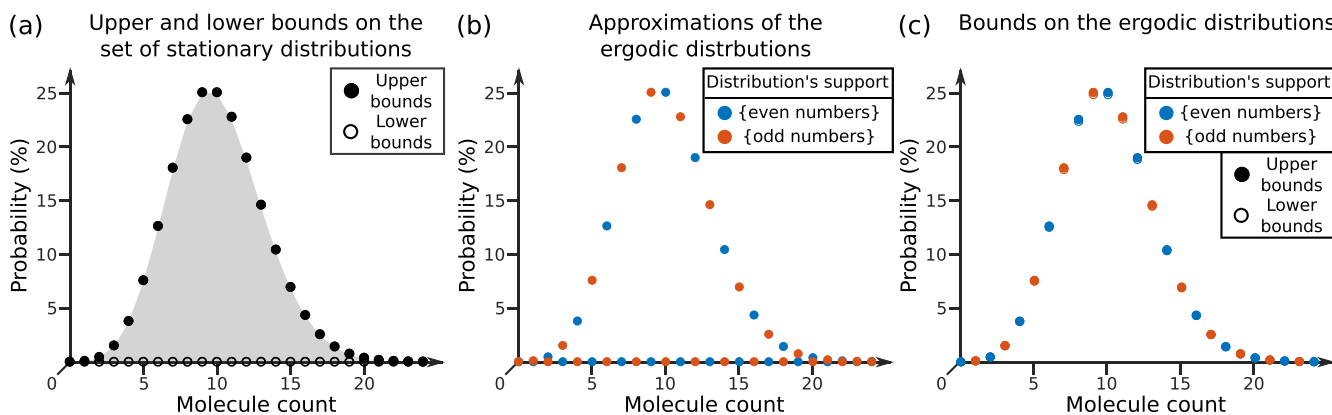


FIG. 8. A dimerization network (116) with multiple stationary distributions. (a) Lower (l^r , open circles) and upper (u^r , filled circles) bounds on the set of stationary distributions for the truncation $\mathcal{S}_{25^7} = \{x^7 < r = 25^7\} = \{0, 1, \dots, 24\}$ (23 equations involving 25 states; solver time = 0.5 s averaging 0.01 s per bound). The gray shading indicates the gap between upper and lower bounds. Note that the lower bounds are zero indicating the presence of multiple stationary distributions. (b) The optimal points $\pi_0^{*,r}$ and $\pi_1^{*,r}$ of the LPs (118) (solver time = 0.01 s per optimal point) provide approximations of the ergodic distributions and indicate that there are two closed communicating classes: the even numbers and the odd numbers. (c) The lower and upper bounds (open and filled circles, visually indistinguishable) computed separately on each of the two ergodic distributions with support on the odd numbers (in red) and even numbers (in blue) (solver time = 0.5 s, average of 0.01 s per bound). Parameter values: $k_1 = 50$ and $k_2 = 0.5$. The lower bounds are now nonzero, indicating the uniqueness of each of the ergodic distributions.

VI. DISCUSSION

We have introduced two mathematical programming approaches that yield bounds on: (i) the stationary moments and (ii) the stationary distributions of biochemical reaction networks. These statistical quantities typically satisfy infinite sets of coupled equations: (i) the stationary moment equations and (ii) the stationary CME. Both our approaches consider a subset of these equations and employ: (i) semidefinite programming and (ii) linear programming to bound the set of solutions. The bounds we obtain provide converging estimates of moments and probabilities with quantifiable errors.

Regarding our first method, which provides bounds for stationary moments, recently, and independently of our work, SDP-based procedures have been proposed by several authors.^{33,35–37} Our work differs from those works in two ways: first, our results apply to networks with both polynomial and *rational* propensities, a wider class of networks of interest in biochemistry, beyond the mass action models considered in Refs. 35–37; second, we give mathematically precise conditions for the validity of the method (Assumption 1) and we explain how these conditions can be verified in practice. To the best of our knowledge, our second approach, the LP bounding and approximation procedure for probability distributions, has not appeared in the CME literature (see Ref. 38 for a discussion of related methodologies in the optimization literature). Importantly, both methods are tightly interlinked: our second method uses SDP moment bounds to formulate the LPs, in order to obtain controlled approximations of stationary solutions and marginals.

For some CMEs, the SDP approach might need to include a large number of moments to obtain accurate estimates of lower order moments (see Figs. 2 and 7). Such large SDPs pose a computational challenge for larger networks as the number of moments $\#_d$ is $\binom{n+d}{d}$, where n denotes the number of species in the network and d is the maximum moment order and thus explodes combinatorially with the number of species. Similar costs are encountered when using moment-closure methods.^{21,22} In contrast with moment closure methods,^{19–22} however, the proposed SDP method to bound moments yields approximations with quantified errors. Furthermore, we show that repeated applications of our SDP method yield upper (respectively, lower) bounds that are monotonically decreasing (respectively, increasing) as the number of moment equations and inequalities is increased (Theorem 5). Although, as mentioned in Sec. III B, there are reaction networks for which the bounds do not converge to the exact moments, they often converge in practice (Fig. 2 and other examples in Refs. 35–37, 44, and 60). In addition, when tight SDP bounds prove computationally too expensive, our LP approach can be used to tighten bounds on moments of interest employing a loose SDP bound on a higher moment (Fig. 7). Finally, it is possible to obtain sharper SDP bounds for restricted state spaces,^{42–44,61} but these refinements are beyond the scope of this paper.

As stated above, the LP approach produces convergent bounds on the stationary solutions (including their marginals or averages). To do so, it uses a moment bound obtained with the SDP method. It is worth remarking that, while we have limited ourselves to rational networks where the moment bound can be obtained using SDPs, the LP approach can be extended beyond rational propensities by

using Foster-Lyapunov criteria.⁴¹ If the CME has a unique solution, our LP method yields converging lower and upper bounds on this solution and easy-to-evaluate error bounds (Corollaries 20, 24, and 25). If the CME has multiple solutions, our method provides bounds over the set of possible solutions. Furthermore, the procedure can be adapted to infer the closed communicating classes and to compute converging approximations of (and bounds on) the corresponding ergodic distributions. Additionally, if we are unsure whether the stationary solution is unique, our method provides a uniqueness test (Corollary 28) that settles the question.

Although LP solvers are highly mature and scalable, the applicability of our LP approach can present computational challenges. First, as discussed in Sec. IV B 3 (step 5), the LPs can become ill-conditioned if the truncation is large although this issue is mitigated by scaling the variables and by ongoing improvements in LP solvers. Second, although the computational cost of solving an LP depends on the algorithm, the cost per bound is at least $\mathcal{O}(|\mathcal{S}_r|)$, where $|\mathcal{S}_r|$ is the size of the truncation. For the purpose of computing the entire distribution, we need $\mathcal{O}(|\mathcal{S}_r|)$ such bounds; hence, the cost is at least $\mathcal{O}(|\mathcal{S}_r|^2)$. If computing a marginal distribution where k species remain, we need $\mathcal{O}(|\mathcal{S}_r|^{k/n})$ bounds with a cost of at least $\mathcal{O}(|\mathcal{S}_r|^{1+k/n})$. If only a stationary average is of interest, the cost is at least $\mathcal{O}(|\mathcal{S}_r|)$ since only two bounds per average need to be computed. Note also that the truncation size typically grows combinatorially in the number of unbounded species, e.g., the number of states for a simplex truncation $\{x \in \mathbb{N}^n : x_1 + \dots + x_n \leq M\}$ is $\binom{n+M}{n}$, where M is an upper cutoff for the species count. Hence, the cost of LPs suffers a combinatorial explosion in the number of species, as for all truncation-based methods.^{28–32,62}

Several other truncation-based schemes have been proposed to approximate the stationary solutions of the CME.^{28–30,32} In contrast with ours, those schemes typically assume that the CME has a unique stationary distribution, which has to be verified separately.^{63,64} Perhaps most extensively studied is the *truncation-and-augmentation* (TA) scheme, originally proposed by Seneta⁶⁵ for discrete-time chains. Its continuous-time counterpart^{28,29,62} converges in total variation for exponentially ergodic chains, monotone chains, and certain generalizations.²⁸ However, bounds on the TA approximation error can be conservative and often involve constants that are difficult to compute in practice.^{28,29,66–72} Spieler *et al.*^{30,31} overcame this issue by iterating the TA scheme and applying a tail bound derived from a Foster-Lyapunov criterion to bound the stationary distribution. Spieler's truncation-based scheme is thus closest to ours. However, their scheme entails solving only systems of linear equations which, although cheaper to compute and simpler to implement than LPs, offer no guarantee of convergence and are only applicable in the unique case.

Another distinct feature of our method is that it enables the direct computation of bounds on the marginal distributions, without the need to compute bounds for each state of the joint distribution. Marginal distributions are of particular interest for the analysis of high-dimensional networks and for inference of model parameters from single cell data. Since our approach yields upper and lower bounds on the marginals, it can be used to bound

the likelihood or likelihood ratios from experimental observations. This would be useful to extend the work in Ref. 73 avoiding error redistribution using water filling methods and aiding by selecting the size of truncations that are sufficient for parameter identifiability. Similar bounds on acceptance ratios could be used in Metropolis-Hastings algorithms to extend the applicability of our method to Bayesian inference. We therefore expect that our approach will be valuable not only for estimating distributions but also for estimating model parameters from noisy single cell data where accurate approximations with quantified errors are needed.

ACKNOWLEDGMENTS

We thank Justine Dattani and Diego Oyarzún for stimulating discussions and Michela Ottobre and Jure Vogrinc for insights on the stability and long-term behavior of continuous-time chains. J.K. gratefully acknowledges support through a BBSRC Ph.D. studentship (Grant No. BB/F017510/1); P.T. through a Fellowship of the Royal Commission for the Exhibition of 1851; G.B.S. through the EPSRC Fellowship Grant No. EP/M002187/1; and M.B. through the EPSRC Grant No. EP/N014529/1 funding the EPSRC Centre for Mathematics of Precision Healthcare.

APPENDIX A: MINIMAL CONTINUOUS-TIME MARKOV CHAINS, THEIR LONG-TERM BEHAVIOR, AND STATIONARY DISTRIBUTIONS

In practice, the chain $X = (X(t))_{t \geq 0}$ in Sec. II is often constructed by running the Gillespie algorithm,^{74–76} i.e., one starts the chain from a state x sampled from an *initial distribution* $\lambda := (\lambda(x))_{x \in \mathcal{S}}$. If $q(x) := -q(x, x)$ in (4) and (5) is zero, leave the chain at x for all time. Otherwise, wait an exponentially distributed amount of time with mean $1/q(x)$, sample $y \neq x$ from the probability distribution $(q(x, y)/q(x))_{y \neq x}$, and update the chain's state to y (we say that the chain *jumps* from x to y , and we call the time at which it jumps the *jump time*). Repeat these steps starting from y instead of x . All random variables sampled must be independent of each other.

If T_n denotes the n th jump time, then the limit

$$T_\infty := \lim_{n \rightarrow \infty} T_n$$

is known as the *explosion time* of the chain, i.e., the first instant by which the chain has left every finite subset of the state space⁷⁷ (Sec. II C). If no such explosion occurs, then $T_\infty = \infty$, and we say that the chain is *nonexplosive*,

$$\mathbb{P}_\lambda(\{T_\infty = \infty\}) = 1, \quad (\text{A1})$$

where \mathbb{P}_λ denotes the probability measure underlying the chain (the subscript λ emphasizes the fact that the starting state was sampled from the distribution λ). If (A1) holds for every probability distribution λ [$\lambda(x) \geq 0, \forall x \in \mathcal{S}, \lambda(\mathcal{S}) = 1$], then the rate matrix Q is *regular*.

We collectively refer to the probabilities

$$(p_t(x))_{x \in \mathcal{S}, t \geq 0} = (\mathbb{P}_\lambda(\{X_t = x, t < T_\infty\}))_{x \in \mathcal{S}, t \geq 0}$$

of observing the process in the state $x := (x_1, \dots, x_n) \in \mathcal{S}$ at time $t \geq 0$ as the *time-varying law* of the chain. The time-varying law

is the minimal non-negative solution of the CME (6) [see Ref. 34 (Corollary 2.21)].

A probability distribution $\pi := (\pi(x))_{x \in \mathcal{S}}$ on \mathcal{S} is said to be a *stationary* (or *steady-state* or *invariant*) *distribution* of the chain if sampling the chain's starting position from π ensures that it will be distributed according to π for all time,

$$\mathbb{P}_\pi(\{X_t = x, t < T_\infty\}) = \pi(x), \quad \forall x \in \mathcal{S}, t \geq 0. \quad (\text{A2})$$

Summing both sides of (A2) over $x \in \mathcal{S}$ and taking the limit $t \rightarrow \infty$, we find that the chain is nonexplosive when its starting location is sampled from a stationary distribution,

$$\mathbb{P}_\pi(\{T_\infty = \infty\}) = 1. \quad (\text{A3})$$

Taking the derivative in time of (A2), we find that stationary distributions are stationary solutions of the CME (6) [that is, it belongs to (9)]. The reverse direction is slightly more complicated:

Theorem 30 (Theorem 2.41³⁴). *Let X be a continuous-time chain with rate matrix Q satisfying (4) and (5). A probability distribution π on \mathcal{S} is a stationary distribution of X if and only if it is a stationary solution of the CME and the chain is nonexplosive when initialized with law π [i.e., (A3) holds].*

In particular, assuming that Q is regular, π is a stationary solution if and only if it is a stationary distribution of the chain. In other words, (9) is an *analytical* (as in nonprobabilistic) *linear programming characterization*^{34,78} of the set of stationary distributions for regular Q . The nonexplosivity in Theorem 30 is crucial: a counterexample is the birth-death process (50) with $a_+(x) := 2^{2x}$ and $a_-(x) := 2^{2x}/2$. In this case, the sum in (55) is finite showing that the CME has a unique stationary solution π given by (54) and (55). However,⁷⁹ (Theorem 11) shows that the process is explosive for any initial distribution (including π) and it follows from (A3) that no stationary distribution exists.

Stationary distributions are of interest because, if the chain is stable, then, regardless of the initial distribution λ , they determine¹⁵ the chain's long term behavior, i.e., the time-varying law of the chain converges in total variation to a stationary distribution π

$$\lim_{t \rightarrow \infty} \|p_t - \pi\| = 0. \quad (\text{A4})$$

Furthermore, the *empirical distribution* ρ_T , which denotes the fraction of the time interval $[0, T]$ that the chain will spend in a state x ,

$$\rho_T(x) := \int_0^{\min\{T, T_\infty\}} 1_x(X_t) dt, \quad \forall x \in \mathcal{S},$$

also converges to π

$$\lim_{T \rightarrow \infty} \|\rho_T - \pi\| = 0 \quad \mathbb{P}_\lambda\text{-almost surely.} \quad (\text{A5})$$

In general, the stationary distributions featuring in (A4) and (A5) depend on the initial distribution λ and on the starting location X_0 , respectively.

If the chain starts in one of the closed communicating classes (defined in Sec. IV B 2), then it can never escape the class.

The convergence of the empirical distribution in (A5) then implies that, in the stable case, there must exist at least one stationary distribution per closed communicating class C_j and that the stationary distribution must have support contained in C_j [$\pi_j(C_j) = 1$]. This distribution π_j is unique, and if the initial distribution λ has its mass contained in C_j , then¹⁵ both the time-varying law and the empirical distribution converge to π_j [in the sense that the π featuring in both (A4) and (A5) is π_j]. For this reason, π_j is known as an *ergodic distribution* of the chain. The definition of the set $\mathcal{T} := \mathcal{S} \setminus \cup_j C_j$ featuring in the decomposition (108) implies that the chain visits any given state x in \mathcal{T} at most finitely many times [in particular, $1_x(X_t) \rightarrow 0$ as $t \rightarrow \infty$ \mathbb{P}_λ -almost surely]. It follows that the chain's paths must eventually either leave \mathcal{T} or diverge to infinity. In the case of a stable chain, tending to infinity is not an option and so the chain eventually enters one of the closed communicating classes. It then follows from (A2) that no stationary distribution π such that $\pi(x) > 0$ exists for a state x in \mathcal{T} . Bringing this discussion together,¹⁵ we have that stationary distribution π in (A5) is the ergodic distribution π_j of the closed communicating class C_j that the chain's path eventually enters, while that in (A4) is a weighted combination of the ergodic distributions where the weight given to π_j is the probability that the chain ever enters C_j . Theorem 27 in Sec. IV B 2 follows from these facts.

APPENDIX B: A FOSTER-LYAPUNOV CRITERION

In practice, verifying whether a chain is stable is done by applying a Foster-Lyapunov criterion. For our examples, we will use the following well-known criterion:

Theorem 31 (Foster-Lyapunov criterion^{15,80,81}). *If there exist constants $K_1 \in \mathbb{R}$, $K_2 > 0$, and a normlike function w (Definition 12) that satisfy*

$$Qw(x) := \sum_{y \in \mathcal{S}} q(x, y)w(y) \leq K_1 - K_2w(x), \quad \forall x \in \mathcal{S},$$

then the following hold:

- (i) The rate matrix Q is regular.
- (ii) There exists at least one stationary distribution.
- (iii) For each closed communicating class, there exists an ergodic distribution with support in that class.
- (iv) Every stationary distribution π satisfies $\langle w \rangle_\pi \leq K_1/K_2 < \infty$.
- (v) The stationary distributions determine the long-term behavior of the chain:
 - (a) for every deterministic starting distribution, there exists a stationary distribution π satisfying (A4);
 - (b) for any starting distribution λ and for \mathbb{P}_λ -almost every path, there exists a stationary distribution π satisfying (A5).

Proof. Part (i) is Ref. 80 (Theorem 1.11) [see also Ref. 81 (Theorem 2.1)]. Parts (ii)–(v) follow from Ref. 15 (Theorems 8.1 and 8.2) and Ref. 81 (Theorem 4.6). \square

We have used this criterion for our examples in the main text, as follows.

1. Schlögl's model

In the case of Schlögl's model (10), fixing $w(x) := x^{d-2}$ for an integer $d > 2$, we have

$$Qw(x) = g_{d-1}(x) - k_2(d-2)x^d,$$

where g_{d-1} is a polynomial of degree $d-1$. Thus,

$$\begin{aligned} Qw(x) &\leq \sup_{x \in \mathbb{N}} \left\{ g_{d-1}(x) - \frac{k_2(d-2)}{2} x^d \right\} - \frac{k_2(d-2)}{2} x^d \\ &\leq \sup_{x \in \mathbb{N}} \left\{ g_{d-1}(x) - \frac{k_2(d-2)}{2} x^d \right\} - \frac{k_2(d-2)}{2} x^{d-2}, \end{aligned}$$

and the supremum is finite. Taking $d \geq 3$, Theorem 31 tells us that (10) has a regular rate matrix and at least one stationary distribution; that the moments $\langle x^l \rangle_\pi, \dots, \langle x^{d-2} \rangle_\pi$ are finite for any stationary distribution π ; and that the limits (A4) and (A5) hold. Since we can choose ever larger d , we have finiteness of all moments. Uniqueness of the distribution follows from (54) and (55).

2. Toggle switch

In the case of the toggle switch chain of Sec. V A, setting $w(x) := (x_1 + x_2)^d$, we have

$$\begin{aligned} Qw(x) &= (a_1(x) + a_3(x))((x_1 + x_2 + 1)^d - (x_1 + x_2)^d) \\ &\quad + (a_2(x) + a_4(x))((x_1 + x_2 - 1)^d - (x_1 + x_2)^d) \\ &\leq (k_1 + k_3)((x_1 + x_2 + 1)^d - (x_1 + x_2)^d) \\ &\quad + (a_2(x) + a_4(x))((x_1 + x_2 - 1)^d - (x_1 + x_2)^d) \\ &\leq g_{d-1}(x_1 + x_2) - d(k_2x_1 + k_4x_2)(x_1 + x_2)^{d-1} \\ &\leq g_{d-1}(x_1 + x_2) - d \min\{k_2, k_4\}(x_1 + x_2)^d, \end{aligned}$$

where g_{d-1} is a polynomial of degree $d-1$. For this reason, proceeding as we did above for Schlögl's model, we have that Q is regular, that the chain does have a stationary distribution, and that all of the moments of each of the stationary distributions are finite. The nontrivial lower bounds in Fig. 4 and Corollary 28 show that it is unique.

3. Bursty gene model

For the bursty gene expression model of Sec. V B, let $w(x) := x_2^d$, $(p(k))_{k \in \mathbb{N}}$ denote the geometric distribution, $p(k) = (1 - p(0))^k p(0)$ with $p(0) = 1/(1 + \langle b \rangle)$, and $\langle b^l \rangle$ denotes its l th moment. We then have

$$\begin{aligned} Qw(x) &= k_4x_1 \sum_{k=0}^{\infty} p(k)((x_2 + k)^d - x_2^d) + k_5x_2((x_2 - 1)^d - x_2^d) \\ &\leq k_4 \sum_{l=0}^{d-1} \binom{d}{l} \langle b^{d-l} \rangle x_2^l + k_5x_2((x_2 - 1)^d - x_2^d) \\ &= g_{d-1}(x_2) - dk_5x_2^d, \end{aligned}$$

where g_{d-1} is a polynomial of degree $d-1$ (note that all moments of a geometric random variable are finite). Because the state space

is $\{0, 1\} \times \mathbb{N}$, w is normlike. Thus, proceeding as we did above for Schlögl's model, we have that Q is regular, that there exists at least one stationary distribution, and that each stationary distribution has all moments finite. For each set of parameter values, we solved LP (90) with $x = (0, 0)$ and obtained a nontrivial lower on $\pi((0, 0))$. Corollary 28 then showed that the stationary distribution is unique.

4. Dimerization network with multiple stationary distributions

For the network (116), choosing $w(x) := x^{d-1}$ with integer $d > 1$, we have

$$Qw(x) = g_{d-1}(x) - 2k_2(d-1)x^d,$$

where g_{d-1} is a polynomial of degree $d - 1$. Proceeding as for Schlögl's model above, we have that Q is regular, that there exists at least one stationary distribution, and that each stationary distribution has all moments finite.

REFERENCES

- M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**, 1183 (2002).
- Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, *Science* **329**, 533 (2010).
- S. Uphoff, N. D. Lord, B. Okumus, L. Potvin-Trottier, D. J. Sherratt, and J. Paulsson, *Science* **351**, 1094 (2016).
- A. Arkin, J. Ross, and H. H. McAdams, *Genetics* **149**, 1633 (1998).
- S. H. Dandach and M. Khammash, *PLoS Comput. Biol.* **6**, e1000985 (2010).
- P. Thomas, N. Popović, and R. Grima, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6994 (2014).
- B. Munsky, Z. Fox, and G. Neuert, *Methods* **85**, 12 (2015).
- F. Fröhlich, P. Thomas, A. Kazerounian, F. J. Theis, R. Grima, and J. Hasenauer, *PLoS Comput. Biol.* **12**, e1005030 (2016).
- G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, *Science* **339**, 584 (2013).
- N. Strelkova and M. Barahona, *J. R. Soc., Interface* **7**, 1071 (2010).
- N. Strelkova and M. Barahona, *Chaos* **21**, 023104 (2011).
- D. A. Oyarzún, J.-B. Lugagne, and G.-B. V. Stan, *ACS Synth. Biol.* **4**, 116 (2014).
- C. Zechner, G. Seelig, M. Rullan, and M. Khammash, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4729 (2016).
- M. Tomazou, M. Barahona, K. M. Polizzi, and G.-B. Stan, *Cell Syst.* **6**, 508 (2018).
- S. P. Meyn and R. L. Tweedie, *Adv. Appl. Probab.* **25**, 487 (1993).
- M. Hemberg and M. Barahona, *Biophys. J.* **93**, 401 (2007).
- M. Hemberg and M. Barahona, *BMC Syst. Biol.* **2**, 42 (2008).
- D. Schnoerr, G. Sanguinetti, and R. Grima, *J. Phys. A* **50**, 093001 (2017).
- S. Engblom, *Appl. Math. Comput.* **180**, 498 (2006).
- S. Engblom and J. Pender, e-print [arXiv:1406.6164](https://arxiv.org/abs/1406.6164) (2014).
- E. Lakatos, A. Ale, P. D. W. Kirk, and M. P. H. Stumpf, *J. Chem. Phys.* **143**, 094107 (2015).
- D. Schnoerr, G. Sanguinetti, and R. Grima, *J. Chem. Phys.* **143**, 185101 (2015).
- N. G. Van Kampen, *Phys. Lett. A* **59**, 333 (1976).
- E. Schwerer, "A linear programming approach to the steady-state analysis of Markov processes," Ph.D. thesis, Stanford University, 1996.
- K. Helmes and R. H. Stockbridge, *Stochastic Models* **19**, 255 (2003).
- O. Hernández-Lerma and J. B. Lasserre, *Markov Chains and Invariant Probabilities* (Birkhäuser, 2003).
- J. Kuntz, M. Ottobre, G.-B. Stan, and M. Barahona, *SIAM J. Sci. Comput.* **38**, A3891 (2016).
- A. G. Hart and R. L. Tweedie, *Appl. Math.* **3**, 2205 (2012).
- A. Gupta, J. Mikelson, and M. Khammash, *J. Chem. Phys.* **147**, 154101 (2017).
- T. Dayar, H. Hermanns, D. Spieler, and V. Wolf, *Numer. Linear Algebra Appl.* **18**, 931 (2011).
- D. Spieler, "Numerical analysis of long-run properties for Markov population models," Ph.D. thesis, Saarland University, 2014.
- T. Dayar, W. Sandmann, D. Spieler, and V. Wolf, *Adv. Appl. Probab.* **43**, 1005 (2011).
- J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona, e-print [arXiv:1702.05468v1](https://arxiv.org/abs/1702.05468v1) (2017).
- J. Kuntz, "Deterministic approximation schemes with computable errors for the distributions of Markov chains," Ph.D. thesis, Imperial College London, 2017.
- Y. Sakurai and Y. Hori, in *IEEE 56th Annual Conference on Decision and Control* (IEEE, 2017), pp. 1206–1211.
- G. R. Dowdy and P. I. Barton, *Comput.-Aided Chem. Eng.* **40**, 2239 (2017).
- K. R. Ghusinga, C. A. Vargas-Garcia, A. Lamperski, and A. Singh, *Phys. Biol.* **14**, 04LT01 (2017).
- J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona, e-print [arXiv:1810.03658](https://arxiv.org/abs/1810.03658) (2018).
- J. R. Norris, *Markov Chains* (Cambridge University Press, 1997).
- F. Schlögl, *Z. Phys.* **253**, 147 (1972).
- P. W. Glynn and A. Zeevi, *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz* (Institute of Mathematical Statistics, 2008), pp. 195–214.
- J. B. Lasserre, *Moments, Positive Polynomials and Their Applications* (Imperial College Press, 2009).
- Semidefinite Optimization and Convex Algebraic Geometry*, edited by G. Blekherman, P. A. Parrilo, and R. R. Thomas (SIAM, 2013).
- G. R. Dowdy and P. I. Barton, *J. Chem. Phys.* **148**, 084106 (2018).
- J. B. Lasserre, *Trans. Am. Math. Soc.* **354**, 631 (2002).
- J. Löfberg, in *IEEE International Symposium on Computer Aided Control System Design* (IEEE, 2004), pp. 284–289.
- M. Nakata, in *IEEE International Symposium on Computer-Aided Control System Design* (IEEE, 2010), pp. 29–34.
- G. Fantuzzi and F. Fuentes, mpYALMIP, version 1.1.2, <https://github.com/giofantuzzi/mpYALMIP>, 2016.
- G. Fantuzzi, D. Goluskin, D. Huang, and S. I. Chernyshenko, *SIAM J. Appl. Dyn. Syst.* **15**, 1962 (2016).
- D. Papp and S. Yildiz, *SIAM J. Optim.* **29**, 822 (2019).
- Y. Zheng, G. Fantuzzi, and A. Papachristodoulou, "Fast ADMM for sum-of-squares programs using partial orthogonality," *IEEE Trans. Autom. Control* (published online 2018).
- I. I. CPLEX, V12.1: User's Manual for CPLEX, 2009.
- C. Gardiner, *Stochastic Methods*, 4th ed. (Springer-Verlag, 2009).
- R. R. Vallabhajosyula, V. Chickarmane, and H. M. Sauro, *Bioinformatics* **22**, 346 (2005).
- See <https://github.com/barahona-research-group/Stationary-bounds-for-continuous-time-chains> for the MATLAB code used to generate the plots presented in Sec. V C.
- T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature* **403**, 339 (2000).
- R. Perez-Carrasco, P. Guerrero, J. Briscoe, and K. M. Page, *PLoS Comput. Biol.* **12**, e1005154 (2016).
- N. Kumar, T. Platini, and R. V. Kulkarni, *Phys. Rev. Lett.* **113**, 268105 (2014).
- V. Shahrezaei and P. S. Swain, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17256 (2008).
- Y. Sakurai and Y. Hori, *J. R. Soc., Interface* **15**, 20170709 (2018).
- M. Laurent, *Emerging Applications of Algebraic Geometry* (Springer-Verlag, 2009), pp. 157–270.
- R. L. Tweedie, *J. Appl. Probab.* **8**, 311 (1971).
- L. Paulevé, G. Craciun, and H. Koepl, *J. Math. Biol.* **69**, 55 (2014).
- A. Gupta and M. Khammash, *SIAM J. Appl. Dyn. Syst.* **17**, 1213 (2018).
- E. Seneta, *Math. Proc. Cambridge Philos. Soc.* **63**, 983 (1967).
- R. L. Tweedie, *J. Appl. Probab.* **35**, 517 (1998).
- S. P. Meyn and R. L. Tweedie, *Ann. Appl. Probab.* **4**, 981 (1994).
- Y. Liu, *Sci. China Math.* **58**, 2633 (2015).

- ⁶⁹H. Masuyama, *J. Oper. Res. Soc. Jpn.* **60**, 271 (2017).
- ⁷⁰H. Masuyama, *Linear Algebra Appl.* **514**, 105 (2017).
- ⁷¹Y. Liu, W. Li, and H. Masuyama, *Oper. Res. Lett.* **46**, 409 (2018).
- ⁷²Y. Liu and W. Li, *Adv. Appl. Probab.* **50**, 645 (2018).
- ⁷³Z. Fox, G. Neuert, and B. Munsky, *J. Chem. Phys.* **145**, 074101 (2016).
- ⁷⁴D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976).
- ⁷⁵D. G. Kendall, *J. R. Stat. Soc. Ser. B* **12**, 116 (1950).
- ⁷⁶W. Feller, *Trans. Am. Math. Soc.* **48**, 488 (1940).
- ⁷⁷J. Kuntz, P. Thomas, G. B. Stan, and M. Barahona, *SIAM J. Sci. Comput.* **41**, A748 (2019).
- ⁷⁸T. G. Kurtz and R. H. Stockbridge, *SIAM J. Control Optim.* **36**, 609 (1998).
- ⁷⁹G. E. H. Reuter, *Acta Math.* **97**, 1 (1957).
- ⁸⁰M. Chen, *J. Appl. Probab.* **28**, 305 (1991).
- ⁸¹S. P. Meyn and R. L. Tweedie, *Adv. Appl. Probab.* **25**, 518 (1993).