Rigorous bounds on the stationary distributions of the chemical master equation via mathematical programming

Juan Kuntz,1,2,a) Philipp Thomas,1,2) Guy-Bart Stan,2,c) and Mauricio Barahona1,4)

1) Department of Mathematics, Imperial College London, London, United Kingdom
2) Department of Bioengineering, Imperial College London, London, United Kingdom

(Dated: 27 June 2019)

The stochastic dynamics of biochemical networks are usually modelled 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 non-unique 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 dimerisation model with multiple stationary distributions.

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 cell1–3. Such reactions are key components in gene regulatory and signalling networks involved in cellular adaptation and cell fate decisions4–6. Mathematically, stochastic reaction networks are modelled 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 inference7,8, both for the identification of molecular mechanisms9 and for the design of synthetic cellular circuits10–14.

Significant effort has been devoted to investigating the stationary solutions of CMEs, which determine the long time behaviour of the stochastic process15. While exact16,17 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 moments19–23. Yet few of these methods provide quantified approximation errors20. 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 schemes28–32 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 work27,33,34, 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 propensities35–37. Here we extend the mathematical framework to rational networks of interest in biochemistry, and we state rigorous, checkable mathematical conditions for the validity of their accuracy...
of the approach.

The second approach employs linear programming to obtain lower and upper bounds on stationary averages of the CME, using state space truncations guided by moment bounds computed with our first (SDP) approach. 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 that converge in total variation to the stationary solution and its marginals with a computable approximation error (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 non-unique case, the scheme provides converging approximations of the ergodic distributions.

The paper is organised as follows. In Section 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 for a simple birth-death process that displays a chemical bifurcation, followed by the general computational approach for multi-species 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 Section 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 Section 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 \( S_1, S_2, \ldots, S_n \):

\[
R_j : v_{ij}^+ S_1 + \cdots + v_{nj}^+ S_n \xrightarrow{a_j} v_{ij}^- S_1 + \cdots + v_{nj}^- S_n \quad (1)
\]

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

Formally, the state of the system is described by the random variable \( X(t) = (X_1(t), \ldots, X_n(t)) \in \mathbb{N}^n \), a vector with components representing the number of molecules of each species at time \( t \). The dynamical process is modelled with a minimal continuous-time Markov chain\(^{39} \) 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 \( v_j := (v_{ij}^+ - v_{ij}^-, \ldots, v_{nj}^+ - v_{nj}^-) \) 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 \( y \):}

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

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

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

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

in which case \( Q \) is said to be totally stable and conservative.

If the Markov chain cannot leave the state space in finite time, the 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)

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

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

\[
p_t Q(x) := \sum_{y \in S} p_t(y)q(y, x), \quad (7)
\]

Following standard convention, probability distributions and measures are defined throughout as row vectors. Any probability distribution \( \pi := (\pi(x))_{x \in S} \) that solves the equation

\[
\pi Q(x) = 0 \quad \forall x \in S, \quad (8)
\]

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

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

of absolutely summable real sequences indexed by states in \( S \). The set of all stationary solutions forms a convex polytope in \( \ell^1 \):

\[
\mathcal{P} := \left\{ \pi \in \ell^1 : \pi Q(x) = 0, \forall x \in S, \quad \pi \geq 0, \sum_{x \in S} \pi(x) = 1 \right\}. \quad (9)
\]
where $\pi \geq 0$ is shorthand for $\pi(x) \geq 0, \forall x \in S$. For most networks of interest the stationary solutions determine the long-term behaviour of the chain (see Appendix A). In many cases, we will be interested in obtaining the $\pi$-average of a real-valued function $f$ on $S$:

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

For example, the $k$-th stationary moment is $\langle x^k \rangle_{\pi}$.

III. BOUNDING THE STATIONARY MOMENTS OF THE CME

As a first use of optimisation 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 distributions of the CME in Sec. IV.

To motivate our optimisation 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 Schlögl\textsuperscript{40} as a model for a chemical bifurcation:

$$2S \xrightleftharpoons{a_1}{a_2} 3S, \quad \varnothing \xrightleftharpoons{a_3}{a_4} S, \quad (10)$$

with mass-action propensities:

$$a_1(x) := k_1x(x-1), \quad a_2(x) := k_2x(x-1)(x-2) \quad (11)$$

$$a_3(x) := k_3, \quad a_4(x) := k_4x, \quad (12)$$

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

The CME of this network has a unique stationary solution $\pi$ 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:

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

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):

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

FIG. 1. Outer approximations and bounds for the moments of the stationary solution of Schlögl’s model (10). (a) Grey 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 colours) 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) (coloured dots) are computed by solving the SDFPs (47)–(48). Parameter values: $k_1 = 1, k_2 = 1, k_3 = 0.8, k_4 = 1$.\[\text{Parameter values: } k_1 = 1, k_2 = 1, k_3 = 0.8, k_4 = 1.\]
where \( b_1 := k_3, b_2 := k_1 + 2k_2 + k_4, b_3 := k_1 + 3k_2, b_4 := k_2 \) are positive numbers. Even after noting that
\[
\langle 1 \rangle_\pi = 1,
\] (15)
as \( \pi \) 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:
\[
\langle x^2 \rangle_\pi - \langle x \rangle^2_\pi \geq 0.
\] (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 \( f := (f_0, f_1)^T \in \mathbb{R}^2 \) 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
\[
\langle f^2 \rangle_\pi = f_0^2 \langle 1 \rangle_\pi + 2 f_0 f_1 \langle x \rangle_\pi + f_1^2 \langle x^2 \rangle_\pi \geq 0
\] (17)
\[
\langle x f^2 \rangle_\pi = f_0^2 \langle x \rangle_\pi + 2 f_0 f_1 \langle x^2 \rangle_\pi + f_1^2 \langle x^3 \rangle_\pi \geq 0
\] (18)
Let us define the vector of moments
\[
z := \left( \langle 1 \rangle_\pi, \langle x \rangle_\pi, \langle x^2 \rangle_\pi, \langle x^3 \rangle_\pi \right)^T \in \mathbb{R}^4.
\]
Then the inequalities (17)–(18) are written compactly as
\[
\langle f^2 \rangle_\pi = f^T M^0_\pi(z)f \geq 0,
\] (19)
\[
\langle x f^2 \rangle_\pi = f^T M^1_\pi(z)f \geq 0,
\] (20)
where the matrices \( M^0_\pi(y) \) and \( M^1_\pi(y) \) are defined by
\[
M^0_\pi(y) := \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix}, \quad M^1_\pi(y) := \begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \end{bmatrix},
\]
for any four-dimensional vector \( y = (y_0, y_1, y_2, y_3)^T \in \mathbb{R}^4 \).
Since (19)–(20) hold for all \( f \in \mathbb{R}^2 \), we have that \( M^0_\pi(z) \) and \( M^1_\pi(z) \) are positive semidefinite (p.s.d.):
\[
M^0_\pi(z) \succeq 0, \quad M^1_\pi(z) \succeq 0.
\] (21)
From Sylvester’s criterion, (21) is equivalent to
\[
z_0 \geq 0, \quad z_1 \geq 0, \quad z_2 \geq 0, \quad z_3 \geq 0,
\] (22)
\[
z_0 z_2 - z_1^2 \geq 0, \quad z_1 z_3 - z_2^2 \geq 0.
\] (23)
\[
z_0 z_2 - z_1^2 \geq 0, \quad z_1 z_3 - z_2^2 \geq 0.
\] (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
\[
\mathcal{E}^3 = \left\{ y \in \mathbb{R}^4 : \begin{array}{l}
y_0 = 1 \\
b_1 y_0 - b_2 y_1 + b_3 y_2 - b_4 y_3 = 0 \\
M^0_\pi(y) \succeq 0 \\
M^1_\pi(y) \succeq 0
\end{array} \right\}.
\] (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:
\[
L^3_\alpha := \inf\{y_\alpha : y \in \mathcal{E}^3\} \leq \langle x^\alpha \rangle_\pi \leq \sup\{y_\alpha : y \in \mathcal{E}^3\} =: U^3_\alpha \\
\alpha = 0, 1, 2, 3.
\] (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
\[
r^+_2(y_1) = y_2 \leq \frac{b_3 x \pm \sqrt{4b_1 b_3 x + (b_3^2 - 4b_2 b_4)x^2}}{2b_4}.
\] (27)
Hence \( \mathcal{E}^3 \) can be rewritten equivalently as:
\[
\mathcal{E}^3 = \left\{ y \in \mathbb{R}^4 : \begin{array}{l}
y_0 = 1 \\
y_1 \geq 0 \\
\max\{y_1^2, r^+_2(y_1)\} \leq y_2 \leq r^+_2(y_1) \\
y_3 = (b_1 - b_2 y_1 + b_3 y_2)/b_4
\end{array} \right\}.
\] (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 \( r^+_2(x) \geq r^+_3(x) \) for all \( x \geq 0 \), the upper bounds in (26) are obtained by the northeasternmost intersection of \( y_2 = 0 \) and \( r^+_2(y_1) \): \((U_1^3, U_2^3) = (r_4, r^+_4(r_4))\), 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:
\[
0 \leq \langle x \rangle_\pi \leq r_4, \quad 0 \leq \langle x^2 \rangle_\pi \leq r^+_2(r_4).
\] (30)
As seen in Fig. 1(a), the analytical bounds based on \( \mathcal{E}^3 \) are rough. However, we show in the following section how to obtain tighter bounds systematically by appending further moment equations and inequalities and solving the associated optimisations over higher order sets \( \mathcal{E}^d \).

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

The approach in the previous section 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
\[
a_j(x) := \frac{b_j(x)}{s(x)} \quad j = 1, \ldots, m,
\] (31)
where \( b_1, \ldots, b_m \) 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: \( x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \), where \( \alpha \) is the multi-index \( (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}^n \) and \( |\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_n \) 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

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

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

\[
   z := (z_\beta)_{|\beta| \leq d}. \quad (36)
\]

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 \( \alpha \)-moment, it is helpful to define the associated polynomial function \( g_\alpha(x) \):

\[
   s(x) Q x^\alpha = \sum_{j=1}^{m} b_j(x) ((x + v_j)^\alpha - x^\alpha) = \sum_{|\beta| \leq d_{g_\alpha}} (g_\alpha)_\beta x^\beta =: g_\alpha(x), \quad (37)
\]

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

\[
   g_\alpha := \left( (g_\alpha)_\beta \right)_{|\beta| \leq d}, \quad (38)
\]

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 \( \alpha \)-moment equation

\[
   \langle Q x^\alpha \rangle_\pi = z^T g_\alpha = 0 \quad (39)
\]

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

**Proof.** Consider the adjoint of (8):

\[
   \langle Q f \rangle_\pi = \sum_{x \in S} \sum_{y \in S} q(x, y) f(y) \pi(x) = 0, \quad (40)
\]

which, by Fubini’s theorem, is valid for any \( f \) as long as \( \sum_{x \in S} |q(x, x) f(x)| \pi(x) \) is finite41. Because

\[
   s(x) |q(x, x) x^\alpha| = \sum_{j=1}^{m} b_j(x) x^\alpha = \sum_{j=1}^{m} \sum_{|\beta| \leq d_b} (b_\beta)_j x^{\alpha+\beta},
\]

then Assumption 1 implies

\[
   \sum_{x \in S} |q(x, x) x^\alpha| \pi(x) \leq \sum_{|\beta| \leq d_b} \left( \sum_{j=1}^{m} (b_\beta)_j \right) z_{\beta+\alpha} < \infty,
\]

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

\[
   0 = \langle Q x^\alpha \rangle_\pi = \left( \frac{g_\alpha}{s} \right)_\pi = \sum_{|\beta| \leq d} (g_\alpha)_\beta \left( \frac{x^\beta}{s} \right)_\pi = z^T g_\alpha. \quad (41)
\]

In addition to the moment equations (39), the moments \( z \) satisfy additional constraints. Firstly, since \( \pi \) is a probability distribution, we have:

\[
   z^T s = \sum_{|\beta| \leq d} s_\beta \left( \frac{x^\beta}{s} \right)_\pi = \left( \frac{s}{s} \right)_\pi = \langle 1 \rangle_\pi = 1. \quad (42)
\]

Furthermore, the rational moments satisfy well-known semidefinite inequalities34,42,43. Specifically, the localising matrices are positive semidefinite:

\[
   M^d_i(z) \succeq 0 \quad \forall i = 0, \ldots, n, \quad (42)
\]

where the \( M^d_i(y) \) are defined by

\[
   [M^d_i(y)]_{\alpha\beta} := y_{\alpha+\beta}, \quad \forall \alpha, \beta : |\alpha|, |\beta| \leq \lfloor d/2 \rfloor,
\]

\[
   [M^d_i(y)]_{\alpha\beta} := y_{\alpha+\beta+e}, \quad \forall \alpha, \beta : |\alpha|, |\beta| \leq \lfloor (d - 1)/2 \rfloor,
\]
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 $|d/2|$ with (column) vector of coefficients $f = (f_\beta | \beta| \leq |d/2|)$, we have

$$f^T M_d^{0}(z) f = \sum_{|\alpha| \leq |d/2|} \sum_{|\beta| \leq |d/2|} f_\alpha f_\beta z_{\alpha + \beta}$$

$$= \left( \sum_{|\beta| \leq |d/2|} f_\beta x^\beta \right) \left( \sum_{|\alpha| \leq |d/2|} f_\alpha x^\alpha \right)$$

$$= \left( \frac{f^2}{s} \right)_\pi \geq 0. \quad (43)$$

Similarly, it can be shown\textsuperscript{42,43} that

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

Since (43)–(44) hold for any vector $f$, the matrices $M_d^{0}(z)$ are positive semidefinite.

Remark 4. In the case of Schlögl’s model (10), we had $s(x) = 1$, $d_0 = 3$ and $\#d = d + 1$. Eq. (14) is the moment equation (39) with $\alpha = 1$, and the matrices in (21) with $d = 3$ are $M_d^{0}(y)$ and $M_d^{1}(y)$.

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

$$\mathcal{E}^d := \left\{ y \in \mathbb{R}^{\#d} : y^T g_\alpha = 0 \quad \forall |\alpha| \leq d - d_0 + 1, \right.$$\n
$$y^T s = 1, \quad M_d^{j}(y) \geq 0 \quad \forall i = 0, 1, \ldots, n. \left\} \right.$$

where $d_0$ is the maximum degree of the numerators in (31), and $s$, $g_\alpha$ and $M_d^{j}(y)$ are defined in (32), (38) and (42), respectively.

Proof. The proof follows from Lemma 3, (41)–(42). \qed

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.

Firstly, 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 (resp. smallest) $\alpha$-entry provides an upper (resp. lower) bound on the $\alpha$-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 (resp. southwesternmost) vector of these outer approximations yield upper (resp. lower) bounds on the first two moments of Schlögl’s model.

Secondly, note that the moment matrix $M_d^{j}(y)$ is a principal submatrix of $M_{d+1}^{j}(y)$. 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).

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

$$L_f^d \leq L_{f+1}^d \leq \cdots \leq L_f^d \leq U_f^d \leq \cdots \leq U_{f+1}^d \leq U_f^d \leq \sum_{|\beta| \leq d} f_\beta \left( \frac{x^\beta}{s} \right)_\pi = z^T f,$$

where $L_f^d := \inf \{ f^T y : y \in \mathcal{E}^d \}$

$$U_f^d := \sup \{ f^T y : y \in \mathcal{E}^d \}. \quad (46)$$

Proof. Note that $z \in \mathcal{E}^d$ because

$$\left( \frac{f}{s} \right)_\pi = \left( \frac{\sum_{|\beta| \leq d} f_\beta x^\beta}{s} \right)_\pi = \sum_{|\beta| \leq d} f_\beta \left( \frac{x^\beta}{s} \right)_\pi = z^T f.$$

Hence (46) follows from (47)–(48). 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 $\mathcal{E}^d$.

Applying these results to $\langle x^\alpha \rangle_\pi$, the $\alpha$-moment of the CME, is straightforward. Let $f(x) := s(x)x^\alpha$ and choose $d \geq |\alpha| + d_0$ 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 (resp. decreasing) sequence of lower (resp. upper) bounds for $\langle x^\alpha \rangle_\pi$:

$$L_{|\alpha|+d_0}^d \leq \cdots \leq L_\alpha^d \leq \langle x^\alpha \rangle_\pi \leq U_\alpha^d \leq \cdots \leq U_{|\alpha|+d_0}^d.$$

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 $d \to \infty$. Although in our experience the bounds often converge numerically, there is no general guarantee for several reasons. Firstly, 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,41 for which more stringent conditions can be produced at a higher computational cost42,44,45.

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)–(48) are the extreme points of the linear functional $y \mapsto f^T y$ over the set $E^d$, which is defined by linear equalities and semidefinite inequalities. Hence, computing the bounds amounts to solving a semidefinite program (SDP), a convex optimisation problem for which there exist efficient computational tools. Therefore, instead of ad hoc analytical manipulations, like those leading to (30), a general procedure by constructing and solving the SDPs is systematically 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 $\pi$ and choose the order of the approximation $d$ for which $d + 1$ stationary moments are finite (Assumption 1) using a Foster-Lyapunov criterion (Theorem 31 in Appendix B).

3. Compute the bounds $L^d_f$ and $U^d_f$ by solving the two SDPs (47)–(48). We set up the SDPs using the modelling package YALMIP46, and solve them using the multi-precision solver SDPA-GMP47 with the interface mpYALMIP48. Examples of computation times are given in the figure captions.

SDPs involving high order moments can be numerically ill-conditioned27,44,49. 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 multi-precision solver SDPA-GMP47 as in Ref.49. Alternatively, one can scale the moments27,44,49, or adapt recently developed specialised solvers50,51.

4. Evaluate the error of the bounds by computing the gap $U^d_f - L^d_f$. 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 case, we recommend breaking the impasse by employing the LP approach of the next section (see Fig. 7(b)).

As an example of this procedure, Fig. 2(a) shows how the computed upper and lower bounds $(L^d_\pi, U^d_\pi)$ for the first three stationary moments of Schlögl’s model become tighter as we increase $d$, the order of the approximation. Fig. 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 the previous section allows us to compute bounds of the stationary moments of the CME by constructing and optimising 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 optimising over outer approximations of the set of stationary solutions of the CME. To do so, we use a moment bound obtained in the previous section. Note that the approximations in this section only involve linear inequalities (instead of semidefinite ones). Hence the optimisations to be solved are linear programs (instead of SDPs), a simpler subclass of convex optimisation 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 $S = \mathbb{N}$ whose value $x \in \mathbb{N}$ changes by $\pm 1$ in each reaction:

\[ \emptyset \xrightarrow{a_{-}} S \xrightarrow{a_{+}} \emptyset. \]  

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

\[ a_{-}(1)\pi(1) - a_{+}(0)\pi(0) = 0, \]

\[ a_{-}(x + 1)\pi(x + 1) - (a_{+}(x) + a_{-}(x))\pi(x) + a_{+}(x - 1)\pi(x - 1) = 0, \quad x \geq 1. \]

Assuming non-vanishing death rates

\[ a_{-}(x) > 0, \quad \forall x \geq 1, \]

it is well known that the unique stationary solution is:

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

(54)

with \(\pi(0)\) given by the normalisation condition:

\[
\pi(0) = \frac{1}{\sum_{x=0}^{\infty} \gamma(x)} =: \frac{1}{\gamma(S)},
\]

(55)

where we have introduced the notation for sums over sets:

\[
\sum_{x \in S} \gamma(x) =: \gamma(S).
\]

(56)

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

1. **Semi-analytical approach for bounds and approximation**

For most birth-death processes, no closed-form expression for \(\gamma(S)\) 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

\[
\mathcal{S}_r := \{x \in \mathbb{N} : x^\alpha < r\} = \{0, 1, \ldots, [r^{1/\alpha}] - 1\},
\]

(57)

with size \(|\mathcal{S}_r| = [r^{1/\alpha}]\) controlled by the parameters \(\alpha \in \mathbb{Z}_+, \text{ and } r > 0\). 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 \(\alpha\)-moment:

\[
(x^\alpha)_x \leq c.
\]

(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

\[
\pi(x) \leq \frac{\gamma(x)}{\gamma(S)} =: u_x \quad \forall x \in \mathcal{S}_r.
\]

(59)

FIG. 2. *Moment bounds for Schlögl’s model (10) using SDPs. (a) Upper bounds \(U^{\alpha}_t\), red circles) and lower bounds \(L^{\alpha}_t\), blue circles) for the first three moments \((x^\alpha)_x\), \(\alpha = 1, 2, 3\) 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, \(d = 3, \ldots, 17\)) for a solver time of 29 seconds (1 second per bound). Parameters: \(k_1 = 6, k_2 = 1/3, k_3 = 50, k_4 = 3\), which correspond to a unimodal stationary solution.
Lower bound: Using (58), we obtain a bound on the probability mass \( m_r \) outside of the truncation:

\[
m_r := \sum_{x \notin S_r} \pi(x) \leq \frac{1}{r} \sum_{x \notin S_r} x^\alpha \pi(x) \leq \frac{(x^n)_r}{r} \leq \frac{c}{r} =: \varepsilon_r,
\]

We say that \( \varepsilon_r \) is a tail bound.

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

\[
\pi(0) = \frac{1 - m_r}{\sum_{x \in S_r} \gamma(x)} \geq 1 - \varepsilon_r
\]

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

\[
\pi(x) \geq u^r_x (1 - \varepsilon_r) =: l^r_x, \quad \forall x \in S_r
\]

Convergent bounds: We have thus shown that

\[
l^r_x \leq \pi(x) \leq u^r_x, \quad \forall x \in S_r,
\]

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

**Approximating the distribution and the approximation error:**

Motivated by these facts, we define the two following measures (lower and upper bounds padded with zeros),

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

\[
u^r := (u^r(x))_{x \in S}, \quad u^r(x) := \begin{cases} u^r_x & \text{if } x \in S_r \\ 0 & \text{if } x \notin S_r \end{cases}
\]

and introduce them as approximations for \( \pi(x) \):

\[
\tilde{\pi} \simeq \pi, \quad \text{where } \tilde{\pi} = l^r \text{ or } \tilde{\pi} = u^r
\]

We quantify the approximation error of \( \tilde{\pi} \) with the total variation norm:

\[
||\pi - \tilde{\pi}|| = \sup_{A \subseteq S} |\pi(A) - \tilde{\pi}(A)|,
\]

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 characterised further. Using (59)–(65), we have:

\[
||\pi - l^r|| = \pi(S) - l^r(S) = 1 - \sum_{x \in S_r} l^r_x
\]

\[
= 1 - (1 - \varepsilon_r) \sum_{x \in S_r} u^r_x = \varepsilon_r,
\]

\[
||\pi - u^r|| = \max \{ u^r(S_r) - \pi(S_r), \pi(S_r) - u^r(S_r) \}
\]

\[
= \max \{ 1 - (1 - m_r), m_r \} = m_r \leq \varepsilon_r.
\]

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

We summarise these findings in the following theorem:

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

(i) The values of the distribution over the truncation are bounded above and below:

\[
l^r \leq \pi(x) \leq u^r, \quad \forall x \in S_r,
\]

where \( u^r = \gamma(x)/\gamma(S_r) \) and \( l^r = u^r \).

(ii) The measures \( l^r = (l^r(x))_{x \in S} \) and \( u^r = (u^r(x))_{x \in S} \) defined in (64)–(65) approximate the solution with approximation errors:

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

(iii) The bounds vary monotonically with \( r \):

\[
l^r \leq l^{r+1} \leq \cdots \leq \pi(x) \leq \cdots \leq u^{r+1} \leq u^r, \quad \forall x \in S_r
\]

and the sequences of approximations converge in total variation to \( \pi \):

\[
\lim_{r \to \infty} ||\pi - l^r|| = \lim_{r \to \infty} ||\pi - u^r|| = 0.
\]

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

**Remark 9.** If Assumption 1 holds, then \( \pi \) satisfies the moment bound (58) with \( c = U^d_\alpha \), where \( \alpha \in \{1, \ldots, d - d_s\} \) and \( U^d_\alpha \) is defined in (48).

**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:

\[
\frac{1}{\pi(0)} = 2F_2 \left( -\frac{c_1 + 1}{2}, \frac{c_1 - 1}{2}; \frac{-c_2 + 1}{2}, \frac{c_2 - 1}{2}; \frac{k_1}{k_2} \right),
\]

where \( 2F_2 \) denotes the generalised hypergeometric function: \( c_1 := 1 - 4k_3/k_1 \) 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 Fig. 3(a)–(b), we compare the approximations \( l^r \) and \( u^r \) given in Theorem 8 (colour shades) to the analytical solution (black lines) of the unimodal and bimodal
2. Reformulation of the bounds as optimisations

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

Consider a truncation \( S_r \) controlled by the parameters \( \alpha, r \in \mathbb{Z}_+ \) with tail bound \( m_r = \pi(S_r^c) \leq c/\epsilon_r \).

**Definition 10** (Restriction of \( \pi \) to \( S_r \)). The restriction of \( \pi \) to \( S_r \) is:

\[
\pi_r(x) := \begin{cases} \pi(x) & \text{if } x \in S_r \\ 0 & \text{if } x \notin S_r \end{cases} \quad \forall x \in S. \tag{70}
\]

The restriction \( \pi_r \) belongs to the convex polytope:

\[
\mathcal{P}^r = \left\{ \pi^r \in \ell^1 : \pi^r \geq 0, \quad 1 - \epsilon_r \leq \pi^r(S) \leq 1 \right\}, \tag{71}
\]

where \( \mathcal{N}_r := \{0, 1, \ldots, \lceil r^{1/\alpha} \rceil - 1\} \), which follows directly from the definition of the restriction (70); the tail bound (60); and the fact that the stationary equations (51)–(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 optimising over \( \mathcal{P}^r \), as stated in the following theorem.

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

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

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

\[
\sup \{ \pi^r(x) : \pi^r \in \mathcal{P}^r \} = \gamma(x)(\sup \{ \pi^r(0) : \pi^r \in \mathcal{P}^r \}) \leq \frac{\gamma(x)}{\gamma(S_r)} = u'(x), \quad \forall x \in S_r, \tag{72}
\]

which follows from \( \pi^r(0) = \pi^r(S_r) / \gamma(S_r) \leq 1 / \gamma(S_r) \). However, \( u' \) clearly satisfies all constraints in (71), including the moment constraint:

\[
c \geq \langle x^\alpha \rangle_u = \frac{\gamma(S_r)}{\gamma(S)} \sum_{x \in S_r} \frac{\gamma(x)}{\gamma(S_r)} x^\alpha + \frac{\gamma(S_r^c)}{\gamma(S)} \sum_{x \notin S_r} \frac{\gamma(x)}{\gamma(S_r^c)} x^\alpha > \frac{\gamma(S_r)}{\gamma(S)} \langle x^\alpha \rangle_u + \frac{\gamma(S_r^c)}{\gamma(S_r^c)} \langle x^\alpha \rangle_u = \langle x^\alpha \rangle_u.
\]
which follows from \( \pi(x) = \gamma(x) / \gamma(S) \) and the inequality
\[
\sum_{x \in S_r} \frac{\gamma(x)}{\gamma(S)} x^\alpha \geq \sum_{x \in S_r} \frac{\gamma(x)}{\gamma(S_r)} x^\alpha = \sum_{x \in S_r} \frac{\gamma(x)}{\gamma(S_r)} r^\alpha > \sum_{x \in S_r} \frac{\gamma(x)}{\gamma(S_r)} x^\alpha = \langle x^\alpha \rangle_{u^r}.
\]

Hence \( u^r \in P_r^r \) and together with (72), this completes the proof.

Importantly, the definition (71) only involves linear equations and inequalities. Therefore optimising over the polytope consists of solving a linear program, a class of optimisations for which there exist powerful computational platforms and algorithms. Furthermore, this optimisation reformulation can be extended seamlessly to arbitrary networks, as expanded in the next section.

**B. Generalisation to arbitrary networks via linear programming**

We now generalise the optimisation 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 \( S \), rate matrix \( Q \) satisfying (4)–(5), and stationary solutions \( \pi \in E^1 \) that form the polytope \( P \) (9).

To characterise the solutions of the CME, we choose a norm-like function \( w \), which plays the same role as the moment function \( (x^\alpha) \) in Section IV A 2.

**Definition 12** (Norm-like function). A function \( w : S \to \mathbb{R} \) is norm-like if it is non-negative
\[
w(x) \geq 0, \quad \forall x \in S
\]
and has finite sublevel sets:
\[
S_r := \{ x \in S : w(x) < r \}. \tag{74}
\]

Furthermore, we require that the growth of \( w \) be dominated by the stationary solution \( \pi \), so that its expectation with respect to \( \pi \) is finite. We summarise 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 \( \pi \), and that we have available a norm-like function \( w \) with sublevel sets \( S_r \) such that every stationary solution \( \pi \) satisfies
\[
\langle w \rangle_{\pi} = \sum_{x \in S} w(x) \pi(x) \leq c, \tag{75}
\]
where \( c \) is a known constant. This inequality can be thought of as a generalisation 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 App. B).

Regarding the moment bound, in the case of networks with rational propensities satisfying Assumption 1, \( w \) can be chosen to be any norm-like 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 criteria\(^{11} \) (see Appendix B).

In analogy with Lyapunov theory, the sublevel sets (74) of \( w \) play an important role in characterising the stationary solutions of the CME. Specifically, we use the sublevel sets \( S_r \) 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 \( S_r \):
\[
m_r := \pi(S_r^c) \leq \frac{1}{r} \sum_{x \in S_r} w(x) \pi(x) \leq \frac{\langle w \rangle_{\pi}}{r} \leq \frac{c}{r} := \varepsilon_r. \tag{76}
\]

Just as in the previous section, this choice yields a sequence of increasing truncations that approach the entire state space:
\[
S_1 \subseteq S_2 \subseteq \ldots, \quad \bigcup_{r=1}^{\infty} S_r = S.
\]

For each truncation, let \( N_r \) denote the set of states \( x \in S_r \) that cannot be reached in a single jump from outside of the truncation:
\[
N_r := \{ x \in S_r : q(z, x) = 0, \quad \forall z \notin S_r \}, \tag{77}
\]
and the associated convex polytope:
\[
P_r := \left\{ \pi' \in E^1 : \begin{array}{l}
\pi'(Q(x)) = 0, \quad \forall x \in N_r \\
\pi'(S_r^c) = 0 \\
1 - \varepsilon_r \leq \pi'(S) \leq 1 \\
\langle w \rangle_{\pi'} \leq c
\end{array} \right\}, \tag{78}
\]
which, analogously to (71), includes all the stationary equations that only involve states in \( S_r \).

We then have the following lemma:

**Lemma 14** (Outer approximations of \( P \)). Suppose that Assumption 13 holds and let \( \pi_{1r} \) be the restriction of \( \pi \) to \( S_r \), as defined in (70). If \( \pi \in P \), then \( \pi_{1r} \in P_r^r \).

**Proof.** This follows directly from (76) and the fact that \( \pi Q(x) = \pi_{1r} Q(x) \), \( \forall x \in N_r \). \( \square \)

The outer approximation property means that optimising over \( P_r^r \) provides convergent bounds on the averages of functions \( f \) on the state space, as summarised in the following theorem.

**Theorem 15** (Convergent bounds of stationary averages). Consider a reaction network (1) with state space
If $\pi \in P$ and $f : S \mapsto \mathbb{R}$ is any real-valued function, then we can bound its averages over the restrictions:

$$l_f^r \leq \langle f \rangle_{\pi_r} \leq u_f^r, \quad \forall r \in \mathbb{Z}_+$$

(79)

where $\pi|_r$ is the restriction of $\pi$ to $S_r$ defined in (70) and the bounds are given by:

$$l_f^r := \inf \{ \langle f \rangle_{\pi^r} : \pi^r \in P^r \}$$

$$u_f^r := \sup \{ \langle f \rangle_{\pi^r} : \pi^r \in P^r \}.$$  

(80)

If we have additional information on $f$, we have the following bounds on the full $\pi$-averages:

(i) If $f(x) \geq 0, \forall x \notin S_r$, then $l_f^r \leq \langle f \rangle_\pi, \forall r \in \mathbb{Z}_+$.

(ii) If $f(x) \leq 0, \forall x \notin S_r$, then $\langle f \rangle_\pi \leq u_f^r, \forall r \in \mathbb{Z}_+$.

(iii) If $\|f\|_\pi < \infty$ (i.e., $f$ is $\pi$-integrable), then

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

(81)

(iv) If the growth of $f$ is strictly dominated by $w$ as the size $r$ of the truncations $S_r$ increases, i.e.,

$$\lim_{r \to \infty} \sup_{x \notin S_r} \frac{|f(x)|}{w(x)} = 0,$$

then $f$ is $\pi$-integrable for all $\pi \in P$ and the sequences of bounds from below ($l_f^r$) and above ($u_f^r$) converge:

$$\lim_{r \to \infty} l_f^r = l_f := \inf \{ \langle f \rangle_\pi : \pi \in P \},$$

$$\lim_{r \to \infty} u_f^r = u_f := \sup \{ \langle f \rangle_\pi : \pi \in P \}.$$  

(83)

Proof. Eq. (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) follows from (79), the moment bound (75), and the following inequality:

$$\sum_{x \notin S_r} |f(x)|\pi(x) \leq \left( \sup_{x \notin S_r} \frac{|f(x)|}{w(x)} \right) \sum_{x \notin S_r} w(x)\pi(x) \leq \left( \sup_{x \notin S_r} \frac{f(x)}{w(x)} \right) \langle w \rangle_\pi.$$  

(84)

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

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

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 $\pi \in P$ (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.

Remark 16 (LP computation). The bounds $l_f^r$ and $u_f^r$ in (80) are obtained by solving two linear programmes (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 $\pi(\pi^r) = l_f^r$ (or $\langle f \rangle_\pi = u_f^r$). The optimal points exist because the LPs are optimisations of a continuous function over a compact non-empty 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 $(\pi^r)_r$ with a quantifiable error given by (i)–(iii); furthermore, under the conditions in (iv), the approximations converge to $(\pi^r)$ 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 $\pi$, i.e.,

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

In this case, the results of Theorem 15 can be strengthened. And in particular, the feasible points $\pi^r \in P^r$ are good approximations of the stationary solution in the sense that they converge to $\pi$ 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 $P = \{ \pi \}$ consists of a single stationary solution $\pi$. Then the upper and lower bounds (80) converge to the average:

$$\lim_{r \to \infty} l_f^r = \lim_{r \to \infty} u_f^r = \langle f \rangle_\pi,$$

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

$$\lim_{r \to \infty} \langle g \rangle_{\pi^r} = \langle g \rangle_\pi$$  

(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.

Remark 18. When $w$ is norm-like, 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 \( \pi^* \) provide arbitrarily accurate approximations to \( \pi \), yet with no quantifiable bound on the approximation error \( ||\pi - \pi^*|| \).

Bounding and approximating a unique stationary solution with quantifiable error:

Corollary 17 can still be used to obtain approximations with quantified errors of the distribution \( \pi \) itself. Specifically, by applying (85) repeatedly with specific \( f \), the indicator function at each point of the truncation. This allows us to compute bounds on each value of the distribution \( \pi(x) \) in \( S_r \).

Given a truncation \( S_r \), let us define the set of indicator functions \( \{1_x\}_{x \in S_r} \), for one for every state in the truncation, where each \( 1_x \) is defined as (3). For each function \( 1_x \) in the set (i.e., for each state in the truncation), we compute the bounds (79) by solving the LPs (80) with \( f = 1_x \). Hence we obtain lower and upper bounds on \( \pi(x) \), for all states in the truncation:

\[
I^l_x := \inf_{x \in S_r} \pi(x), \quad I^u_x := \sup_{x \in S_r} \pi(x) \quad \forall x \in S_r \tag{87}
\]

where \( I^l_x = I^u_x = \inf\{\pi^f(x) : \pi^f \in \mathcal{P}^r\} \tag{88} \)

\[
u^l_x = u^l_x = \sup\{\pi^f(x) : \pi^f \in \mathcal{P}^r\}. \tag{89}
\]

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

\[
I^l := (I^l(x))_{x \in S}, \quad I^u(x) := \begin{cases} I^l_x & \text{if } x \in S_r \\ 0 & \text{if } x \notin S_r \end{cases} \tag{90}
\]

\[
u^l := (\nu^l(x))_{x \in S}, \quad u^u(x) := \begin{cases} \nu^l_x & \text{if } x \in S_r \\ 0 & \text{if } x \notin S_r \end{cases} \tag{91}
\]

These approximations of \( \pi \) have controlled erros, as summarised 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 \( I^l \) (90) and \( u^u \) (91) fulfill the following:

(i) The approximations \( I^l \) and \( u^u \) bound \( \pi \) from below and above, respectively,

\[
I^l(x) \leq \pi(x) \leq u^u(x) \quad \forall x \in S_r \tag{92}
\]

\[
\pi(x) \leq u^u(x) \quad \forall x \in S_r \tag{93}
\]

with quantified approximation errors \( \varepsilon^l_r \) and \( \varepsilon^u_r \):

\[
||I^l - \pi|| = 1 - I^l(S_r) = \varepsilon^l_r, \tag{94}
\]

\[
||u^u - \pi|| = \max\{u^u(S_r) - 1 + m_r, m_r\} \leq \max\{u^u(S_r) - 1 + c/r, c/r\} = \varepsilon^u_r, \tag{95}
\]

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

(ii) As the truncation size \( r \) approaches infinity (and \( S_r \) approaches \( S \)), the approximation \( u^u \) converges pointwise to the unique \( \pi \),

\[
\lim_{r \to \infty} u^u(x) = \pi(x), \quad \forall x \in S \tag{96}
\]

and the approximation \( I^l \) converges to \( \pi \) in weak* (86) to \( \pi \).

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

\[
|u^u(A) - \pi(A)| \leq \max\{u^u(S_r) - \pi(S_r), |u^u(S_r) - \pi(S_r)|\}, \quad \forall A \subseteq S. \tag{97}
\]

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

\[
|\pi(x) - I^l(x)| \leq \sum_{x \in S_r} |f(x)| (\pi(x) - I^l(x))
+ 2 \sum_{x \notin S_r} |f(x)| \pi(x).
\]

Using the pointwise convergence of \( I^l \) 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 \( I^l \) follows. See Theorem 4.1 in Ref. 38 for details.

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

\[
\lim_{r \to \infty} \varepsilon^l_r = 0,
\]

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

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

To instead answer the question ‘when is \( r \) too small?’ (i.e., to establish how large the partition \( S_r \) 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 \( I^l \) (90) and \( u^u \) (91) cannot be made smaller than the tail bound or the mass of the tail, respectively, i.e.,

\[
||I^l - \pi|| \geq \varepsilon_r, \quad \forall r \in \mathbb{Z}_+ \tag{97}
\]

\[
||u^u - \pi|| \geq m_r, \quad \forall r \in \mathbb{Z}_+, \tag{98}
\]

where \( m_r = \pi(S_r) \leq \varepsilon_r = c/r \).

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

\[
\frac{1 - \varepsilon_r}{\pi^{*,r}(S_r)} \pi^{*,r} \in \mathcal{P}^r,
\]
which implies \( \pi^{*\pi}(S_r) = 1 - \varepsilon_r \) (due to the minimality of \( \pi^{*\pi}(x) \)). Since \( \ell' \) bounds from below all feasible points of \( P_r \), it follows that \( \ell'(S_r) \leq \pi^{*\pi}(S_r) = 1 - \varepsilon_r \). 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)–(98) are sharp for birth-death processes with \( w(x) = x^\alpha \) (see (67)–(68)).

**Approximating marginal distributions:**

For high-dimensional state spaces, we are often interested in marginal distributions rather than the full multivariate solution \( \pi \) defined on \( S \). A marginalisation is associated with a partition of the state space into a collection of disjoint subsets:

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

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

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

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

Typically, we are interested in marginalising the distribution of a reaction network with \( n \) species and state space \( S = \mathbb{N}^n \) 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:

\[ \{A_i\}_{i \in \mathbb{N}} \text{ where } A_i := \mathbb{N}^{k-1} \times \{i\} \times \mathbb{N}^{n-k}, \]

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:

\[ \ell_i^* = \inf \{ \pi^*(A_i) : \pi^* \in P^* \}, \quad \forall i \in \mathcal{I}_r \]

\[ \hat{\ell}_r(\hat{\pi}(i)) := \begin{cases} \ell_i^* & \text{if } i \in \mathcal{I}_r \\ 0 & \text{if } i \notin \mathcal{I}_r \end{cases}, \]

\[ u_r^* = \sup \{ \pi^*(A_i) : \pi^* \in P^* \}, \quad \forall i \in \mathcal{I}_r \]

\[ \hat{u}_r(\hat{\pi}(i)) := \begin{cases} u_i^* & \text{if } i \in \mathcal{I}_r \\ 0 & \text{if } i \notin \mathcal{I}_r \end{cases}, \]

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

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

\[ \ell_r := (\ell_r(\hat{\pi}(i)))_{i \in \mathcal{I}}, \quad \ell_r := \begin{cases} \ell_i^* & \text{if } i \in \mathcal{I}_r \\ 0 & \text{if } i \notin \mathcal{I}_r \end{cases}, \]

\[ \hat{\ell}_r(\hat{\pi}(i)) := \begin{cases} \ell_i^* & \text{if } i \in \mathcal{I}_r \\ 0 & \text{if } i \notin \mathcal{I}_r \end{cases}, \]

which are the analogues for the marginal distribution of the approximations to the entire distribution (90)–(91), and have similar (but not identical) properties, as summarised 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 \( S_r \), then the associated marginal distribution \( \hat{\pi} \) is lower bounded by the approximation \( \hat{\ell}_r \) defined in (103):

\[ \hat{\ell}_r(\hat{\pi}(i)) = 1 - \hat{\ell}_r(\mathcal{I}_r) =: \hat{\ell}_r^r. \]

Furthermore, as the truncation size is increased \( (r \to \infty) \) and \( S_r \) approaches \( S \), \( \hat{\ell}_r \) 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 \( g(x, y) = 1_{A_{\infty}}(x) \)) 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

\[ ||\hat{u}_r - \hat{\pi}|| \leq \max \{ u_i^*(\mathcal{I}_r) - 1 + m_r, m_r \} \]

\[ \leq \max \{ \hat{u}_r(\mathcal{I}_r) - 1 + c/r, c/r \} =: \hat{\varepsilon}_r^u. \]

Furthermore, as the truncation \( S_r \) approaches \( S \), \( \hat{u}_r \) converges pointwise to \( \hat{\pi} \):

\[ \lim_{r \to \infty} \hat{u}_r(i) = \hat{\pi}(i), \quad \forall i \in \mathcal{I}. \]

**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 \( \hat{u}_r \) bounds the marginal \( \hat{\pi}(i) \) if and only if \( A_i \cap S_r^c = \emptyset \), i.e., when the set \( A_i \) is fully contained inside the truncation \( S_r \). Hence \( \hat{u}_r(i) \) does not provide an upper bound if the truncation does not include all the space of the marginalised variables.

However, using the fact that the probability mass of \( A_i \cap S_r^c \) is bounded by the mass of the tail \( m_r \) (76), we have the following easy (but loose) upper bounds:

\[ \hat{\pi}(i) \leq \hat{u}_r(i) + c/r \quad \forall i \in \mathcal{I}_r. \]
2. Non-uniqueness, ergodic distributions and a uniqueness test

Theorem 15 shows that our LP optimisation over the polytopes $P^r$ yields bounds on the stationary averages, even if there are multiple stationary solutions. In the non-unique case, however, the gap between the lower bounds and the upper bounds will reflect the fact that the extreme points of $\pi \rightarrow f_\pi$ over $P$ can be achieved by different solutions in the polytope. Yet it is possible to characterise 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

$$\emptyset \rightarrow 2S_1 \rightarrow \emptyset, \quad S_2 \rightarrow \emptyset,$$

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

$$S = \{(x_1,0) : x_1 \in \mathbb{N} \text{ is odd}\} \cup \{(x_1,0) : x_1 \in \mathbb{N} \text{ is even}\} \cup \{(x_1,x_2) : x_1, x_2 \in \mathbb{Z}_+\} = C_1 \cup C_2 \cup T,$$

where $C_1$ and $C_2$ are closed communicating classes and $T$ contains the remaining states. A set $C \subseteq S$ is a closed communicating class if the chain can transit between any pair of states in $C$ but cannot leave $C$. Another common source of multiple closed communicating classes are conservation laws in reaction networks. For instance, the reactions

$$2S_1 \rightleftharpoons S_2,$$

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 summarised in the following theorem that compiles some facts that are broadly known in the literature.

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

$$S = (\cup_j C_j) \cup T,$$

where $C_j$ are closed communicating classes and $T$ contains the remaining states.

(i) For each $C_j$, there is at most one stationary solution $\pi_j$; hence $\pi_j(C_j) = 1$. Whenever it exists, $\pi_j$ is known as the ergodic distribution associated with $C_j$.

(ii) Let $J$ be the set of indexes $j$ of the ergodic distributions $\pi_j$. The set of stationary solutions $P$ (9) is the set of convex combinations of the ergodic distributions:

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

**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. (Th. 3.5.2)), and (ii) is given in Ref. (Th. 3.4).

Theorem 27 states that the ergodic distributions $\pi_j$ are orthogonal to each other and that they are the extreme points of the convex polytope $P$ (9) of stationary solutions of the CME. Because $P$ is contained in the non-negative orthant of $\ell^1$, 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 summarised in the following corollary.

**Corollary 28** (A uniqueness test). If Assumption 13 holds and $Q$ is regular, then $P = \{\pi\} \iff \exists x \in S, \, r \geq 1 : l^r(x) > 0$.

**Proof.** The proof is by contradiction. Suppose that $P$ is not a singleton. Theorem 27(ii) implies that $P$ contains two or more ergodic distributions, $\pi_j$, each associated with a different closed communicating class, $C_j$. Let us consider, e.g., 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 $l^r$ (Theorem 15(ii)). Hence, $P$ must be a singleton. The converse follows from the convergence of the bounds in Corollary 20(ii).

In other words, if $\pi$ is unique, then the lower bound of any (and all) states in the support of $\pi$ is non-zero for sufficiently large $r$. Conversely, finding a single non-zero lower bound for any $x \in S$ 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 $S$.

When $\pi$ is unique, $l^r$ or $u^r$ are good approximations of the stationary solution. However, this is not so in the non-unique case. Indeed, it is easy to show that, in the non-unique case, the lower and upper bounds are always loose: Corollary 28 shows that the lower bounds are trivially zero everywhere ($l^r = 0$), 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:

$$\liminf_{r \to \infty} u^r(S) \geq |J|, \quad (110)$$

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 $\pi^r$. To this end, we require the following generalisation of Corollary 17.
Corollary 29 (Convergent approximations of ergodic distributions). Let $f$ be any function that satisfies (82) (i.e., $f$ is dominated by the norm-like function $w$ as the size $r$ of the truncations increases), and let $(\pi^{*,r})_{r \in \mathbb{Z}_+}$ be a sequence of optimal points such that
\[ \pi^{*,r} \in \mathcal{P}^r \quad \text{and} \quad \langle f \rangle_{\pi^{*,r}} = u^r = \sup\{\langle f \rangle_{\pi^r} : \pi^r \in \mathcal{P}^r\}, \]
for all $r \in \mathbb{Z}_+$.

(i) If there exists a unique point $\pi^* \in \mathcal{P}$ such that
\[ \langle f \rangle_{\pi^*} = \sup\{\langle f \rangle_{\pi^r} : \pi^r \in \mathcal{P}^r\} \quad (111) \]
then the sequence of optimal points $\pi^{*,r}$ converges to $\pi^*$ in weak* as $r \to \infty$.

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

Proof. The proof of (i) is similar to that of Theorem 15(iv). See Corollary 3.6(ii) and Remark 3.7 in Ref. 39 for details. (ii) follows from (i), Theorem 27(ii), and the fact that indicator functions satisfy (82). \qed

Corollary 29 provides a rationale for how to use our computational framework to obtain approximations of the ergodic distributions $\pi_j$ in the non-unique 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 $\mathcal{C}_j$ with ergodic distribution $\pi_j$, Corollary 29(ii) shows that $\pi^r$ will converge to $\pi_j$ 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 $\mathcal{C}_j$ (if any) the state belongs. Once the class is known, we replace $\mathcal{S}$ with $\mathcal{C}_j$ and proceed as for the unique case to obtain bounds on $\pi_j$. See Section 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 $\epsilon$, we proceed as follows:

1. Verify the existence of stationary solutions $\pi$ and the finiteness of their moments (Assumption 1) using a Foster-Lyapunov criterion (Theorem 31 in App. B).
2. Choose a norm-like 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 $\alpha$ induce smaller truncation sizes $|\mathcal{S}_r| \approx \sqrt{r}$.

To guide the selection of $w$, one can run the scheme with various $w$ to gain information about the shape of the distribution.

3. 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 YALMIP46, SDPA-GMP47, and mpYALMIP48 to formulate and solve the SDP (48) with $w := w$. See details in Sec. III B 2.

4. Choose an initial truncation size $r$ based on the achievable errors established in Proposition 22. To guarantee an error smaller than our tolerance $\epsilon$, we must choose an initial $r > c/\epsilon$.

5. 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 V12.6.32 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.

6. Evaluate the error of the approximations $l_r$ and $u_r$ using (94) and (95), respectively. If the error is larger than our tolerance $\epsilon$, we increase the truncation size $r$ and return to the previous step.

7. 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 5–6:

- If we want to approximate a marginal distribution, we solve the LPs (101)–(102) and quantify the error using (105)–(106).
- If we are interested in a particular stationary average $\langle f \rangle_{\pi^r}$, we instead solve the LPs (80) and control the error using the bounds in Theorem 15(i)–(iii).

8. As the particular reaction network could have several stationary solutions, we check in Step 5 for non-trivial 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 3.5GHz processor and 16GB 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 non-unique case with a dimerisation network. The code used to compute the approximations and bounds for this last example is available at.

A. A toggle switch

Toggle switches are common motifs in many cell-fate decision genetic circuits\(^{6,56,57}\). A simple such circuit consists of two mutually repressing genes\(^{56}\). In particular, we consider the asymmetric case with mutual repression modelled via Hill functions and dilution/degradation modelled via linear decay:

\[
\emptyset \xrightarrow{a_1} P_1 \xrightarrow{a_2} \emptyset, \quad \emptyset \xrightarrow{a_3} P_2 \xrightarrow{a_4} \emptyset. \tag{112}
\]

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:

\[
a_1(x) = \frac{k_1}{1 + (x_2/\theta)^3}, \quad a_2(x) = k_2 x_1, \\
a_3(x) = \frac{k_3}{1 + x_1}, \quad a_4(x) = k_4 x_2. \tag{113}
\]

where the \( k_i > 0 \) are kinetic constants and \( \theta > 0 \) is the dissociation constant of \( P_1 \).

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

We pick the norm-like 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 \( = 3.6 \) minutes).

We then solve the LPs (88)–(89) and compute the bounding approximations \( \hat{u}^r \) and \( \hat{l}^r \) of the stationary solutions. The fact that the lower bounds \( \hat{l}^r \) are non-zero 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 \( 2.6 \times 10^{-3} \), as given by (94)–(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. Fig. 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 24–25).

Finally, we apply the method to chart the change of the stationary solution as a function of a parameter. In particular, the dissociation constant of protein \( P_2 (\theta) \) can be thought of as a bifurcation parameter: increasing \( \theta \) allows for higher expression of protein \( P_1 \). Fig. 6 presents the lower bounds \( \hat{l} \) on the marginals of both proteins. At small values of \( \theta \), we observe a single population with high numbers of \( P_2 \) repressing \( P_1 \). At large values of \( \theta \), the opposite happens: the population we observe has high numbers of \( P_1 \) repressing \( P_2 \). For intermediate \( \theta \), 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 model\(^{58}\) 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 variable\(^{50}\) with mean \( \langle b \rangle \). The protein represses its own production by switching off the promoter:

\[
\begin{align*}
G_{off} & \xrightarrow{a_1} G_{on}, \\
G_{on} + P & \xrightarrow{a_2} G_{off} + P, \\
G_{on} & \xrightarrow{a_4} G_{on} + b \, P, \\
P & \xrightarrow{a_5} \emptyset.
\end{align*} \tag{114}
\]
we increase the number of moment equations in our 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, we 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 \rangle = 4.5 \times 10^{13} \text{ (time=1 minute per bound)}. We then set $w(x) := x_2^6$ and $c := U_{x_2}^{13}$ in Theorem 15 to obtain:

$$l_{x_2} = \langle x_2 \rangle \leq u_{x_2} + \frac{c}{\sqrt[7/2]{\varepsilon}}.$$  

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

\[
A_i = \{0, 1\} \times \{i\}, \quad \forall i \in \mathbb{N} : i^6 < r =: I_r,
\]

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

FIG. 5. Marginal distributions of the toggle switch (112). (a) Tight lower bounds \(\tilde{u}^l\) 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 seconds, average of 0.17 seconds per bound). For comparison, we show simulations performed using the Gillespie Algorithm with \(10^6\) samples (black dots). (b) Controlled approximations \(\tilde{u}^l\) for increasing state space truncations \(r = 30^6\) (cyan), \(40^6\) (purple), and \(60^6\) (pink). (c) Approximation error \(\tilde{e}^i\) of the lower bound \(\tilde{u}^l\) (red) and bound \(\tilde{e}^i\) on error of \(\tilde{u}^r\) (black). In total, 3900 bounds were computed (solver time = 48.8 minutes, average of 0.75 seconds per bound). Parameters as in Fig. 4.

C. Dimerisation network with multiple stationary distributions

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

\[
\emptyset \xrightarrow{a_1} 2S \xrightarrow{a_2} \emptyset,
\]

with state space \(S = \mathbb{N}\) 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 \(C_0\) and the odd numbers \(C_1\) are closed communicating classes. Furthermore, because the network obeys detailed balance\(^23\), it is straightforward to obtain analytical expressions for the two ergodic distributions:

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

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

To illustrate the use of our tools in such a non-unique 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 (App. B).

Next, we obtain bounds on the stationary solutions. Using the norm-like 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 seconds) to obtain the moment bound

\[ \langle w \rangle \leq c = 4.8814 \times 10^7. \]

We then solve the LPs (88)–(89) to compute upper and lower bounds, \( w^r \) and \( l^r \) (Fig. 8(a)). Note that the lower bounds remain trapped at zero, indicating that the stationary solution is non-unique (as given by Theorem 28). Furthermore, we observe that the mass of the upper bounds approaches two as the truncation grows (i.e., \( w^r(S_r) \to 2 \) as \( r \to \infty \)), 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

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

approach distributions with support on the even and odd numbers, respectively (Fig. 8(b)). These observations in-
dicate that the communicating classes are the even numbers $C_0$ and the odd numbers $C_1$ (Corollary 29).

To verify this claim, we compute two different sets of upper and lower bounds over $C_0$ and $C_1$ separately. When computed over the subsets $C_0$ and $C_1$ separately, the lower bounds are non-zero, 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 $8 \times 10^{-3}$, while that of the upper bounds (95) is bounded above by $8 \times 10^{-3}$. 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.

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: firstly, 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.\(^{35–37}\); secondly, 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 optimisation 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 (resp. lower) bounds that are monotonically decreasing (resp. 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.\(^{35–37,44,60}\)). 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). Lastly, 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. Firstly, 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. Secondly, although the computational cost of solving an LP depends on the algorithm, the cost per bound is at least $O(|S_r|)$, where $|S_r|$ is the size of the truncation. For the purpose of computing the entire distribution, we need $O(|S_r|)$ such bounds; hence the cost is at least $O(|S_r|^2)$. If computing a marginal distribution where $k$ species remain, we need $O(|S_r|^{1+k/n})$ bounds with an cost of at least $O(|S_r|^{1+k/n})$. If only a stationary average is of interest, the cost is at least $O(|S_r|)$, 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 $\{x \in \mathbb{N}^n: x_1 + \cdots + x_n \leq M\}$ is $\binom{n+M}{n}$, where $M$ is an upper cut-off 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 pro-
posed to approximate the stationary solutions of the CME.\textsuperscript{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.\textsuperscript{63,64} Perhaps most extensively studied is the truncation-and-augmentation (TA) scheme, originally proposed by Seneta\textsuperscript{65} for discrete-time chains. Its continuous-time counterpart\textsuperscript{28,29,62} converges in total variation for exponentially ergodic chains, monotone chains, and certain generalisations.\textsuperscript{28} However, bounds on the TA approximation error can be conservative and often involve constants that are difficult to compute in practice.\textsuperscript{28,29,66–72} Spieler et al.\textsuperscript{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 bounds on the TA approximations with quantified errors.

VII. ACKNOWLEDGEMENTS

We thank Justine Dattani and Diego Oyarzún for stimulating discussions, and Michela Ottobre and Jure Vogrinc for insights on the stability and long-term behaviour of continuous-time chains. JK gratefully acknowledges support through a BBSRC PhD studentship (BB/F017510/1); PT through a Fellowship of the Royal Commission for the Exhibition of 1851; GBS through the EPSRC Fellowship EP/M002187/1; and MB through the EPSRC grant EP/N014529/1 funding the EPSRC Centre for Mathematics of Precision Healthcare.

Appendix A: Minimal continuous-time Markov chains, their long-term behaviour and stationary distributions

In practice, the chain $X = (X(t))_{t \geq 0}$ in Sec. II is often constructed by running the Gillespie Algorithm,\textsuperscript{74–76} i.e., one starts the chain from a state $x$ sampled from an initial distribution $\lambda := (\lambda(x))_{x \in S}$. If $q(x) := -q(x,x)$ in (4)–(5) is zero, leave the chain at the $x$ 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

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

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 space (Sec. 2.3). If no such explosion occurs, then $T_\infty = \infty$, and we say that the chain is non-explosive:

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

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

We collectively refer to the probabilities

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

of observing the process in the state $x := (x_1, \ldots, x_n) \in S$ 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.\textsuperscript{34} (Cor. 2.21)).

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

$$\mathbb{P}_\pi \{X_t = x, t < T_\infty\} = \pi(x), \quad \forall x \in S, \quad t \geq 0.$$  \hspace{1cm} (A2)

Summing both sides of (A2) over $x \in S$ and taking the limit $t \to \infty$, we find that the chain is non-explosive when its starting location is sampled from a stationary distribution:

$$\mathbb{P}_\pi \{T_\infty = \infty\} = 1.$$  \hspace{1cm} (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:

\begin{itemize}
  \item[a)] \textbf{Appendix A: Minimal continuous-time Markov chains, their long-term behaviour and stationary distributions}
  \item[b)] In practice, the chain $X = (X(t))_{t \geq 0}$ in Sec. II is often constructed by running the Gillespie Algorithm,\textsuperscript{74–76} i.e., one starts the chain from a state $x$ sampled from an initial distribution $\lambda := (\lambda(x))_{x \in S}$. If $q(x) := -q(x,x)$ in (4)–(5) is zero, leave the chain at the $x$ 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.
  \item[c)] If $T_n$ denotes the $n$th jump time, then the limit
  \begin{equation}
  T_\infty := \lim_{n \to \infty} T_n
  \end{equation}
  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 space (Sec. 2.3). If no such explosion occurs, then $T_\infty = \infty$, and we say that the chain is non-explosive:
  \begin{equation}
  \mathbb{P}_\lambda \{T_\infty = \infty\} = 1,
  \end{equation}
  where $\mathbb{P}_\lambda$ denotes the probability measure underlying the chain (the subscript $\lambda$ emphasises the fact that the starting state was sampled from the distribution $\lambda$). If (A1) holds for every probability distribution $\lambda (\lambda(x) \geq 0, \forall x \in S, \lambda(S) = 1)$, then the rate matrix $Q$ is regular.
  \item[d)] We collectively refer to the probabilities
  \begin{equation}
  (p_t(x))_{x \in S, t \geq 0} = (\mathbb{P}_\lambda \{X_t = x, t < T_\infty\})_{x \in S, t \geq 0}
  \end{equation}
  of observing the process in the state $x := (x_1, \ldots, x_n) \in S$ 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.\textsuperscript{34} (Cor. 2.21)).
  \item[e)] A probability distribution $\pi := \{\pi(x)\}_{x \in S}$ on $S$ is said to be a stationary (or steady-state or invariant) distribution of the chain if sampling the chain’s starting position from $\pi$ ensures that it will be distributed according to $\pi$ for all time:
  \begin{equation}
  \mathbb{P}_\pi \{X_t = x, t < T_\infty\} = \pi(x), \quad \forall x \in S, \quad t \geq 0.
  \end{equation}
  Summing both sides of (A2) over $x \in S$ and taking the limit $t \to \infty$, we find that the chain is non-explosive when its starting location is sampled from a stationary distribution:
  \begin{equation}
  \mathbb{P}_\pi \{T_\infty = \infty\} = 1.
  \end{equation}
  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:
\end{itemize}
Theorem 30 (Theorem 2.4115). Let $X$ be a continuous-time chain with rate matrix $Q$ satisfying (4)–(5). A probability distribution $\pi$ 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 non-explosive when initialised with law $\pi$ (i.e., (A3) holds).

In particular, assuming that $Q$ is regular, $\pi$ is a stationary distribution if and only if it is a stationary distribution of the chain. In other words, (9) is an analytical (as in non-probabilistic) linear programming characterisation15,79 of the set of stationary distributions for regular $Q$. The non-explosivity in Theorem 30 is crucial: a counterexample is the birth-death process (50) with $a_+(x) := 2^x$ and $a_-(x) := 2^x/2$. In this case, the sum in (55) is finite showing that the CME has a unique stationary solution $\pi$ given by (54)–(55). However,79 (Theorem 11) shows that the process is explosive for any initial distribution (including $\pi$) 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 $\lambda$, they determine15 the chain’s long term behaviour. I.e., the time-varying law of the chain converges in total variation to a stationary distribution, $\pi$:

$$\lim_{t \to \infty} \|p_t - \pi\| = 0. \tag{A4}$$

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

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

also converges to $\rho$:

$$\lim_{T \to \infty} \|\rho_T - \pi\| = 0 \quad \mathbb{P}_\lambda-\text{almost surely}, \tag{A5}$$

In general, the stationary distributions featuring in (A4) and (A5) depend on the initial distribution $\lambda$ 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 $\pi_j$ is unique and, if the initial distribution $\lambda$ has its mass contained in $C_j$, then15 both the time-varying law and the empirical distribution converge to $\pi_j$ (in the sense that the $\pi$ featuring in both (A4) and (A5) is $\pi_j$). For this reason, $\pi_j$ 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 $\mathbf{1}_x(X_t) \to 0$ as $t \to \infty$ $\mathbb{P}_\lambda$-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 $\pi$ such that $\pi(x) > 0$ exists for a state $x$ in $\mathcal{T}$. Bringing this discussion together15, we have that stationary distribution $\pi$ in (A5) is the ergodic distribution $\pi_j$ 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 $\pi_j$ 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 criterion15,50,61).** If there exist constants $K_1 \in \mathbb{R}$, $K_2 > 0$, and a norm-like function $w$ (Definition 12) that satisfy

$$Qw(x) := \sum_{y \in \mathcal{S}} q(x,y)w(y) \leq K_1 - K_2 w(x) \quad \forall x \in \mathcal{S},$$

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 $\pi$ satisfies $\langle w \rangle_\pi \leq K_1/K - 2 < \infty$.

(v) The stationary distributions determine the long-term behaviour of the chain:

a) for every deterministic starting distribution there exists a stationary distribution $\pi$ satisfying (A4);

b) for any starting distribution $\lambda$ and for $\mathbb{P}_\lambda$-almost every path, there exists a stationary distribution $\pi$ satisfying (A5).

**Proof.** Part (i) is Ref.50 (Theorem 1.11) (see also Ref.61 (Theorem 2.1)). Parts (ii)–(v) follow from Ref.15 (Theorems 8.1 and 8.2) and Ref.81 (Theorem 4.6). □

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

$$\text{Proof.}$$
a. **Schlögl’s model**: In the case of Schlögl’s model (10), fixing \( w(x) := x^{d-2} \) for an integer \( d > 2 \), we have

\[
Q w(x) = g_{d-1}(x) - k_2(d-2)x^d,
\]

where \( g_{d-1} \) is a polynomial of degree \( d-1 \). Thus,

\[
Q w(x) \leq \sup_{x \in \mathbb{N}} \left\{ g_{d-1}(x) - \frac{k_2(d-2)}{2} x^d \right\} - \frac{k_2(d-2)}{2} x^d
\]

\[
\leq \sup_{x \in \mathbb{N}} \left\{ g_{d-1}(x) - \frac{k_2(d-2)}{2} x^d \right\} - \frac{k_2(d-2)}{2} x^{d-2},
\]

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 \( \langle x^1 \rangle_\pi, \ldots, \langle x^{d-2} \rangle_\pi \) are finite for any stationary distribution \( \pi \); and that the limits (A4)–(A5) hold. Since we can choose ever larger \( d \), we have finiteness of all moments. Uniqueness of the distribution follows from (54)–(55).

b. **Toggle switch**: In the case of the toggle switch chain of Sec. VA, setting \( w(x) := (x_1 + x_2)^d \), we have

\[
Q w(x) = (a_1(x) + a_3(x))((x_1 + x_2 + 1)^d - (x_1 + x_2)^d)
\]

\[+ (a_2(x) + a_4(x))((x_1 + x_2 - 1)^d - (x_1 + x_2)^d)
\]

\[\leq (k_1 + k_3)((x_1 + x_2 + 1)^d - (x_1 + x_2)^d)
\]

\[+ (a_2(x) + a_4(x))((x_1 + x_2 - 1)^d - (x_1 + x_2)^d)
\]

\[\leq g_{d-1}(x_1 + x_2) - d(k_2x_1 + k_4x_2)(x_1 + x_2)^{d-1}
\]

\[\leq g_{d-1}(x_1 + x_2) - d \min\{k_2, k_4\}(x_1 + x_2)^d,
\]

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 are finite of each of the stationary distributions are finite. The non-trivial lower bounds in Fig. 4 and Corollary 28 show that it is unique.

c. **Bursty gene model**: For the bursty gene expression model of Sec. VB, let \( w(x) := x_2^2 \) and \( p(k) \) denotes the geometric distribution, \( p(k) = (1-p(0))^k p(0) \) with \( p(0) = 1/(1+\langle b \rangle) \), and \( \langle b \rangle \) denotes its \( l^{th} \) moment. We then have

\[
Q w(x) = k_4x_1 \sum_{k=0}^{\infty} p(k)((x_2 + k)^d - x_2^d)
\]

\[+ k_5x_2((x_2 - 1)^d - x_2^d)
\]

\[\leq k_4 \sum_{k=0}^{\infty} \left\{ \langle b \rangle^k \right\} (x_2 - 1)^d + k_5x_2((x_2 - 1)^d - x_2^d)
\]

\[= g_{d-1}(x_2) - d k_5x_2,
\]

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 norm-like. 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 non-trivial lower on \( \pi(0,0) \). Corollary 28 then showed that the stationary distribution is unique.

d. **Dimerisation network with multiple stationary distributions**: For the network (116), choosing \( w(x) := x^{d-1} \) with integer \( d > 1 \), we have

\[
Q w(x) = g_{d-1}(x) - 2k_2(d-1)x^d,
\]

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**

1. M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science 297, 1183 (2002).
2. Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science 329, 533 (2010).
3. S. Uphoff, N. D. Lord, B. Okumus, L. Potvin-Trottier, D. J. Sherratt, and J. Paulsson, Science 351, 1094 (2016).
4. A. Arkin, J. Ross, and H. H. McAdams, Genetics 149, 1633 (1998).
5. S. H. Dandach and M. Khammash, PLOS Comput. Biol. 6, e1000985 (2010).
6. P. Thomas, N. Popović, and R. Grima, Proc. Natl. Acad. Sci. 111, 6994 (2014).
7. B. Minsky, Z. Fox, and G. Neuert, Methods 85, 12 (2015).
8. F. Fröhlich, P. Thomas, A. Kazeroonian, F. J. Theis, R. Grima, and J. Hasenauer, PLoS Comput. Biol. 12, e1005030 (2016).
9. G. Neuert, B. Minsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, Science 339, 584 (2013).
10. N. Strelkowa and M. Barahona, Journal of the Royal Society Interface 7, 1071 (2010).
11. N. Strelkowa and M. Barahona, CHAOS 21, 104 (2011).
12. D. A. Oyarzun, J.-B. Lugagne, and G.-B. V. Stan, ACS Synth. Biol. 4, 116 (2014).
13. C. Ze che r, G. Seelig, M. Rullan, and M. Kham m ash, Proc. Natl. Acad. Sci. 113, 4729 (2016).
14. M. Tomazou, M. Barahona, K. M. Polizzi, and G.-B. Stan, Cell Systems 6, 508 (2018).
15. S. P. Meyn and R. L. Tweedie, Adv. in Appl. Probab. 25, 487 (1993).
16. M. Hemberg and M. Barahona, Biophys. J. 93, 401 (2007).
17. M. Hemberg and M. Barahona, BMC Syst. Biol. 2, 42 (2008).
18. D. Schnoerr, G. Sanguinetti, and R. Grima, J. Phys. A 50, 093001 (2017).
19. S. Engblom, Appl. Math. Comput. 180, 498 (2006).
20. S. Engblom and J. Pender, (2014), arXiv:1406.6164.
21. E. Lakatos, A. Ale, P. P. W. Kirk, and M. P. H. Stumpf, J. Chem. Phys. 143, 94107 (2015).
22. D. Schnoerr, G. Sanguinetti, and R. Grima, J. Chem. Phys. 143, 185101 (2015).
23. N. G. Van Kampen, Phys. Lett. A 59, 333 (1976).
24. E. Schwemer, A linear programming approach to the steady-state analysis of Markov processes, Ph.D. thesis, Stanford University (1996).
25. K. Helms and R. H. Stockbridge, Stoch. Model. 19, 255 (2003).
26. O. Hernández-Lerma and J. B. Lasserre, Markov chains and invariant probabilities (Birkhäuser, 2003).
27. J. Kunz, M. Ottobre, G.-B. Stan, and M. Barahona, SIAM J. Sci. Comput. 38, A3891 (2016).
28. A. G. Hart and R. L. Tweedie, Appl. Math. 3, 2205 (2012).
29. A. Gupta, J. Mikkelson, and M. Khammash, J. Chem. Phys. 147, 154101 (2017).
30. T. Dayar, H. Hermans, D. Spieler, and V. Wolf, Numer. Linear Algebra Appl. 18, 931 (2011).
