



# Tighter Bounds on Transient Moments of Stochastic Chemical Systems

Flemming Holtorf<sup>1</sup> · Paul I. Barton<sup>1</sup>

Accepted: 21 October 2023 / Published online: 21 November 2023  
© The Author(s) 2023

## Abstract

The use of approximate solution techniques for the Chemical Master Equation is a common practice for the analysis of stochastic chemical systems. Despite their widespread use, however, many such techniques rely on unverifiable assumptions and only a few provide mechanisms to control the approximation error quantitatively. Addressing this gap, Dowdy and Barton (J Chem Phys 149(7):074103, 2018) proposed an optimization-based technique for the computation of guaranteed bounds on the moment trajectories associated with stochastic chemical systems, thereby providing a general framework for rigorous uncertainty quantification. Here, we present an extension of this method. The key contribution is a new hierarchy of convex necessary moment conditions that build upon partitioning of the time domain. These conditions reflect the temporal causality that is inherent to the moment trajectories associated with stochastic processes described by the Chemical Master Equation and can be strengthened by simple refinement of the time domain partition. Analogous to the original method, these conditions generate a hierarchy of semidefinite programs that furnishes monotonically improving bounds on the trajectories of the moments and related statistics of stochastic chemical systems. Compared to its predecessor, the presented hierarchy produces bounds that are at least as tight and features new bound tightening mechanisms such as refinement of the time domain partition which often enable the computation of dramatically tighter bounds with lower computational cost. We analyze the properties of the presented hierarchy, discuss some aspects of its practical implementation and demonstrate its merits with several examples.

---

Communicated by Benoît Chachuat.

---

✉ Paul I. Barton  
pib@mit.edu  
Flemming Holtorf  
holtorf@mit.edu

<sup>1</sup> Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

**Keywords** Semidefinite programming · Polynomial optimization · Moment-sum-of-squares hierarchy · Optimal control · Stochastic chemical kinetics · Stochastic reaction networks · Biochemical systems

## 1 Introduction

The analysis of systems undergoing chemical reactions lies at the heart of many scientific and engineering activities. While deterministic models have proved adequate for the analysis of systems at the macroscopic scale, they often fall short for meso- and microscopic systems, in particular for those that feature low molecular counts. In this regime, the complex and chaotic motion of molecules reacting upon collision causes effectively stochastic fluctuations of the molecular counts that are large compared to the mean and, as a consequence, can have a profound effect on the system's characteristics—a situation frequently encountered in cellular biology [7, 8, 24, 49]. In the context of the continuously growing capabilities of synthetic biology, this fact motivates the use of stochastic models for the identification, design and control of biochemical reaction networks. However, while in these applications stochastic models provide the essential fidelity relative to their deterministic counterparts, their analysis is generally more involved, often requiring heuristic approximations and simplifications. In this article, we address this gap by presenting a convex optimization-based framework for the rigorous quantification of uncertainty in stochastic reaction networks.

Stochastic chemical systems are canonically modeled as jump processes; the system state, as encoded by the molecular counts of the individual chemical species, is modeled to change discretely in response to reaction events as triggered by the arrivals of Poisson processes whose rates depend on the underlying reaction mechanism. The Chemical Master Equation (CME) is an ordinary differential equation that describes how the probability distribution of the state of such a process evolves over time. Specifically, a solution of the CME tracks the probability to observe the system in any reachable state over time. This introduces a major challenge for the analysis of stochastic reaction networks in practice as such systems routinely feature millions or even infinitely many reachable states, rendering a direct solution of the CME intractable. As a consequence, sampling techniques such as Gillespie's Stochastic Simulation Algorithm [27, 28] have become the most prominent approach for the analysis of systems described by the CME. And although sampling techniques perform remarkably well across a wide range of problems, they are inadequate in certain settings. Most notably, they do not scale well for stiff systems, generally do not provide hard error bounds and the evaluation of sensitivity information is challenging [30]. In particular, the two latter shortcomings limit their utility in the context of identification, design and control. Approaches based on finite state projection [50] come with guaranteed error bounds and straightforward sensitivity evaluation, however, generally suffer severely from a large number of reachable states.

From a practical perspective, stochastic reaction networks are often sufficiently characterized by only a few low-order moments of the distribution of the state, for example through means and variances. In that case, tractability of the CME may be

recovered by solving for the moments of its solution directly. The dynamics of a finite sequence of moments associated with the distribution described by the CME, however, generally do not form a closed system of differential equations and hence do not admit a solution by simple (numerical) integration. Numerous moment closure approximations [5, 40, 52, 67] have been proposed to remedy this problem. A major shortcoming of moment closure approximations, however, is that they generally rely on unverifiable assumptions and therefore introduce an uncontrolled error. In fact, it is well-known that their application can lead to unphysical results such as spurious oscillations and negative mean molecular counts, especially when applied to systems with low molecular counts [31, 63, 64].

In order to address the shortcomings of moment closure approximations while preserving the advantages of the moment-based description, several authors have recently proposed schemes for the computation of theoretically guaranteed bounds for the moments (or related statistics) associated with stochastic reaction networks involving low molecular counts; such bounding schemes have been proposed for stationary [21, 25, 43, 59, 62], transient [19, 22, 61] and exit time distributions [10] of stochastic chemical systems and analogous techniques have been successfully applied for the study of other types of stochastic processes [32, 33, 35, 39, 42, 48]. The key insight underpinning all these bounding schemes is that the moment-sum-of-squares hierarchy [45, 54] generates a rich set of convex conic conditions characterizing the true moments associated with stochastic chemical systems in many practically relevant settings. This insight enables the use of conic optimization to approximate the true moment sequence by one that conforms with a finite subset of these “necessary moment conditions” while maximizing (or minimizing) a moment of interest to bound its true value. By imposing necessary moment conditions involving higher and higher-order moments, a sequence of monotonically improving bounds of the true moment of interest is generated.

The unique ability to rigorously quantify errors and uncertainty makes bounding schemes particularly attractive for tasks such as robustness analysis and the verification of approximation techniques like moment closure approximations. The utility of moment bounds for such tasks, however, is directly tied to the bound quality. And while for stationary moments the existing bounding schemes are found to produce remarkably good bounds in a wide range of cases [21, 25, 43, 59, 60], bounds on transient moments are often lacking in tightness, even for simple reaction networks [22, 61]. For transient problems, necessary moment conditions involving moments of rather high order need to be considered to obtain informative bounds. This imposes strong practical limitations as the bounding schemes suffer from the curse of dimensionality, i.e., the size of the bounding problems grows combinatorially with the order of moments considered and the number of chemical species in the system. These limitations are exacerbated by the notorious numerical ill-conditioning of moment problems [10, 20, 21]. In order to address this issue, we extend in this work the bounding scheme proposed by Dowdy and Barton [22] for transient moments of the solutions of the CME. We devise a moment bounding scheme involving a new hierarchy of necessary moment conditions which is motivated by the characteristics of stochastic chemical systems and not generally considered in the standard moment-sum-of-squares hierarchy. In broad strokes, these conditions arise from a partitioning of the time domain in a way that

is akin to discretization techniques commonly used for solving differential equations. As such, they reflect the temporal causality that is inherent to solutions of the CME. These conditions give rise to new bound tightening mechanisms beyond increasing the order of moments considered. For example, a simple refinement of the time domain partition results in a strengthening of the conditions and thus acts as a bound tightening mechanism. Critically, these bound tightening mechanisms avoid augmentation of the order of moments considered and thus enjoy favorable scaling properties in addition to alleviating numerical difficulties associated with the consideration of high-order moments. While these mechanisms do not obviate increasing the moment truncation order and so do not avoid the curse of dimensionality entirely, we find that they greatly improve upon the practicality of Dowdy and Barton's [22] proposal. Cibulka et al. [17] recently studied a closely related spatio-temporal partitioning approach in the context of overapproximating the region of attraction of deterministic control systems via sum-of-squares techniques [36] and report similar improvements in terms of practicality. While their approach is similar in spirit to ours, it takes a different (dual) perspective of approximating continuous functions by piecewise polynomials as opposed to the moment-centered view presented here. Moreover, their approach is geared toward deterministic control systems and does not apply without modification to bounding the moment trajectories of stochastic chemical systems.

The remainder of this article is organized as follows. In Sect. 2, we introduce definitions and assumptions, formally define the problem of bounding the transient moments of stochastic chemical systems and review essential preliminaries. Section 3 is devoted to the development and analysis of the proposed hierarchy of necessary moment conditions. In Sect. 4, we discuss several aspects pertaining to the use of these conditions for computation of moment bounds in practice. The utility of the resultant bounding scheme is demonstrated with several examples in Sect. 5 before we conclude with some open questions in Sect. 6.

## 2 Preliminaries

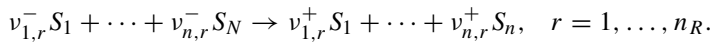
### 2.1 Notation

We denote scalars with lowercase symbols without emphasis, while vectors and matrices are denoted by bold lower- and uppercase symbols, respectively. Throughout, vectors are assumed to be column vectors. Generic sets are denoted by uppercase symbols without emphasis. For special or commonly used sets, we use the standard notation; for example, the (non-negative)  $n$ -dimensional reals and integers are denoted by  $\mathbb{R}^n$  ( $\mathbb{R}_+^n$ ) and  $\mathbb{Z}^n$  ( $\mathbb{Z}_+^n$ ), respectively. Similarly, we refer to the set of symmetric and symmetric positive semidefinite (psd)  $n$ -by- $n$  matrices with  $\mathbb{S}^n$  and  $\mathbb{S}_+^n$ , respectively, and use the usual shorthand notation  $\mathbf{A} \succeq \mathbf{B}$  for  $\mathbf{A} - \mathbf{B} \in \mathbb{S}_+^n$ . The set of  $n$ -dimensional vector and symmetric matrix polynomials with real coefficients (of degree at most  $k$ ) in the variables  $\mathbf{x} = [x_1 \ \dots \ x_N]^\top$  will be denoted by  $\mathbb{R}^n[\mathbf{x}]$  ( $\mathbb{R}_k^n[\mathbf{x}]$ ) and  $\mathbb{S}^n[\mathbf{x}]$  ( $\mathbb{S}_k^n[\mathbf{x}]$ ), respectively. In order to concisely denote multivariate monomials, we employ the multi-index notation: for a monomial in  $n$  variables corresponding to the multi-index

$\mathbf{j} = [j_1 \dots j_n]^\top \in \mathbb{Z}_+^n$ , we write  $\mathbf{x}^{\mathbf{j}} = \prod_{i=1}^n x_i^{j_i}$ . The indicator function of a set  $A$  is denoted by  $\mathbb{1}_A$ . Lastly, we denote the set of  $n$  times continuously differentiable functions on an interval  $I \subset \mathbb{R}$  by  $\mathcal{C}^n(I)$  while the set of absolutely continuous functions is denoted by  $\mathcal{AC}(I)$ . The remaining symbols will be defined as they are introduced.

## 2.2 Problem Statement, Definitions & Assumptions

We consider a reaction system featuring  $n$  chemical species  $S_1, \dots, S_n$  undergoing  $n_R$  different reactions. The system state  $\mathbf{x}$  is encoded by the molecular counts of the individual species, i.e.,  $\mathbf{x} = [x_1 \dots x_n]^\top \in \mathbb{Z}_+^n$ . It changes in response to reaction events according to a given stoichiometry:



In other words, the system state changes by  $\mathbf{v}_r = [v_{1,r}^+ - v_{1,r}^- \dots v_{n,r}^+ - v_{n,r}^-]^\top \in \mathbb{Z}^n$  upon the event of reaction  $r$ . We will restrict ourselves to the framework of stochastic chemical kinetics for modeling such systems.

The notion of stochastic chemical kinetics treats the position and velocities of all molecules in the system as random variables; reactions are assumed to occur at collisions with a prescribed probability. Consequently, the evolution of the system state is a continuous-time jump process. Here, we will assume that this jump process can be described by the Chemical Master Equation (CME).

**Assumption 1** Let  $P_\pi(\mathbf{x}, t)$  be the probability to observe the system in state  $\mathbf{x}$  at time  $t$  given the distribution  $\pi$  of the initial state of the system. Then,  $P_\pi(\mathbf{x}, t)$  satisfies

$$\frac{\partial P_\pi}{\partial t}(\mathbf{x}, t) = \sum_{r=1}^{n_R} a_r(\mathbf{x} - \mathbf{v}_r) P_\pi(\mathbf{x} - \mathbf{v}_r, t) - a_r(\mathbf{x}) P_\pi(\mathbf{x}, t), \quad P_\pi(\cdot, 0) = \pi, \quad (\text{CME})$$

where  $a_r$  denotes the propensity of reaction  $r$ , i.e.,  $a_r(\mathbf{x})dt$  quantifies the probability that reaction  $r$  occurs in  $[0, dt)$  as  $dt \rightarrow 0$  given the initial system state is  $\mathbf{x}$ .

Moreover, we will restrict our considerations to the case of polynomial reaction propensities.

**Assumption 2** The reaction propensities  $a_r$  in (CME) are polynomials.

To ensure the moment trajectories remain well-defined at all times, we will further assume that the stochastic process is well-behaved in the following sense.

**Assumption 3** The number of reaction events occurring in the system within finite time is finite with probability 1.

A consequence of Assumption 3 is that the continuous-time jump process associated with (CME) is regular [57], i.e., it does not explode in finite time. We wish to emphasize

that Assumptions 1–3 are rather weak; Assumptions 1 and 2 are in line with widely accepted microscopic models [29], while Assumption 3 should intuitively be satisfied for any practically relevant system for which the CME is a reasonable modeling approach. Furthermore, Assumption 3 is formally necessary for (CME) to be valid on an indefinite time horizon [57]. For a detailed, physically motivated derivation of the CME alongside discussion of the underlying assumptions and potential relaxations thereof, the interested reader is referred to Gillespie [29].

Instead of studying the probability distribution  $P_\pi$  as a description of the system behavior, in this paper we will focus on its moments defined as follows.

**Definition 2.1** Let  $X$  be the reachable set of the system, i.e.,  $X = \{\mathbf{x} \in \mathbb{Z}_+^n \mid \exists t \geq 0 : P_\pi(\mathbf{x}, t) > 0\}$ , and  $\mathbf{j} \in \mathbb{Z}_+^n$  be a multi-index. The  $\mathbf{j}$ th moment of  $P_\pi(\cdot, t)$  is defined as  $y_{\mathbf{j}}(t) = \sum_{\mathbf{x} \in X} \mathbf{x}^{\mathbf{j}} P_\pi(\mathbf{x}, t)$ .  $y_{\mathbf{j}}$  is said to be of order  $|\mathbf{j}| = \sum_{i=1}^n j_i$ . The function  $y_{\mathbf{j}}(\cdot)$  is called the trajectory of the  $\mathbf{j}$ th moment or the  $\mathbf{j}$ th transient moment. If  $y_{\mathbf{j}}$  further converges to a stationary value in the limit of long times, we refer to that value as the corresponding stationary moment.

Additionally, it will prove useful to introduce the following notion of generalized moments.

**Definition 2.2** Let  $y_{\mathbf{j}}$  be as in Definition 2.1 and  $t_T > 0$ . Consider a uniformly bounded Lebesgue integrable function  $g : [0, t_T] \rightarrow \mathbb{R}$ . The  $\mathbf{j}$ th generalized moment of  $P_\pi$  with respect to  $g$  is defined by  $z_{\mathbf{j}}(g; t) = \int_0^t g(\tau) y_{\mathbf{j}}(\tau) d\tau$  for  $t \in [0, t_T]$ . We say  $g$  is a test function and generates  $z_{\mathbf{j}}(g; t)$ .

Under Assumptions 1 and 2, it is well-known that the dynamics of the  $\mathbf{j}$ th moment are described by a linear time-invariant ordinary differential equation (ODE) of the form

$$\frac{dy_{\mathbf{j}}}{dt}(t) = \sum_{|\mathbf{k}| \leq |\mathbf{j}|+q} c_{\mathbf{k}} y_{\mathbf{k}}(t) = \mathbf{c}^\top \mathbf{y}(t), \tag{1}$$

where  $q = \max_{1 \leq r \leq n_R} \text{deg}(a_r) - 1$ . The coefficient vector  $\mathbf{c}$  can be readily computed from the reaction propensities and stoichiometry; see for example Gillespie [26] for details. For  $q > 0$ , it is clear from Eq. (1) that the dynamics of moments of a certain order in general depend on moments of a higher order. This issue is commonly termed the *moment closure problem*. If we denote by  $\mathbf{y}_L$  the vector of “lower-order” moments up to a specified order, say  $m$ , and by  $\mathbf{y}_H$  the vector of “higher-order” moments of order  $m + 1$  to  $m + q$ , it is clear from Eq. (1) that we obtain a linear time-invariant ODE system of the form

$$\frac{d\mathbf{y}_L}{dt}(t) = \mathbf{A}_L \mathbf{y}_L(t) + \mathbf{A}_H \mathbf{y}_H(t)$$

with  $\mathbf{A}_L \in \mathbb{R}^{n_L \times n_L}$  and  $\mathbf{A}_H \in \mathbb{R}^{n_L \times n_H}$ , where  $n_L = \binom{n+m}{n}$  and  $n_H = \binom{n+m+q}{n} - n_L$  denote the number of lower- and higher-order moments, respectively. For the sake of

a more concise notation, throughout we will often omit these subscripts and instead write

$$\mathbf{K} \frac{d\mathbf{y}}{dt}(t) = \mathbf{A}\mathbf{y}(t), \quad (\text{mCME})$$

where  $\mathbf{A} = [\mathbf{A}_L \ \mathbf{A}_H]$ ,  $\mathbf{K} = [\mathbf{I}_{n_L \times n_L} \ \mathbf{0}_{n_L \times n_H}]$  and  $\mathbf{y} = [\mathbf{y}_L^\top \ \mathbf{y}_H^\top]^\top$ .

In the presence of the moment closure problem, it is clear from the setup of Equation (mCME) that it does not provide sufficient information to determine uniquely the moment trajectories associated with the solution of (CME). In the following, we therefore address the question of how to compute hard, theoretically guaranteed bounds on the true moment trajectory  $y_j(\cdot)$  associated with the solution of (CME) in this setting. To that end, we build on the work of Dowdy and Barton [22] who have recently proposed an approach to answer this question. In broad strokes, they generate upper and lower bounds by optimizing a moment sequence truncated at a given order subject to a set of *necessary moment conditions*, i.e., conditions that the true moment trajectories are guaranteed to satisfy. By increasing the truncation order, the bounds can be improved. Our contribution is an extension of Dowdy and Barton's work in the form of a hierarchy of new necessary moment conditions based upon partitioning of the time domain of the problem. We show that these conditions provide additional, more scalable bound tightening mechanisms beyond increasing the truncation order.

## 2.3 Necessary Moment Conditions

The bounding method proposed by Dowdy and Barton [22] hinges on necessary moment conditions which restrict the set of potential solutions of Eq. (mCME) as much as possible while remaining computationally tractable. Necessary moment conditions in the form of affine constraints and linear matrix inequalities (LMI) have proved to fit that bill [38, 46]. Conditions of this form are of particular practical value as they allow for the computation of the desired bounds via semidefinite programming (SDP). In fact, analogous conditions to those proposed by Dowdy and Barton [22] have been employed for a wide range of applications concerned with the moment-based study of dynamical systems [38], for example for optimal control [37, 47], region of attraction computation [17, 36], the analysis of PDEs [11, 41] or options pricing [48]. In general, affine moment conditions arise from the system dynamics, while the LMIs reflect constraints on the support of the underlying probability distributions. In the following, we will sketch the derivation of these conditions and highlight key properties which will be leveraged for the construction of new necessary moment conditions in Sect. 3.

### 2.3.1 Linear Matrix Inequalities

The fact that the solution of the CME,  $P_\pi(\cdot, t)$ , is a non-negative measure on  $\mathbb{R}^n$  and supported only on the reachable set  $X$  of the underlying reaction system implies that its truncated moment sequences satisfy certain LMIs [21, 25, 43, 45, 59]. The following

argument reveals this fact: Consider a polynomial  $f \in \mathbb{R}[\mathbf{x}]$  that is non-negative on  $X$ ; further, let  $\mathbf{b}$  be a vector polynomial obtained by arranging the elements of the monomial basis of the polynomials up to degree  $d = \lfloor \frac{m+q-\text{deg}(f)}{2} \rfloor$  in a vector. The following generalized inequality

$$\mathbb{E}[f\mathbf{b}\mathbf{b}^\top] \succeq P_\pi(\hat{\mathbf{x}}, t) f(\hat{\mathbf{x}}) \mathbf{b}(\hat{\mathbf{x}}) \mathbf{b}(\hat{\mathbf{x}})^\top \succeq \mathbf{0}, \quad \forall(\hat{\mathbf{x}}, t) \in \mathbb{R}^n \times \mathbb{R}_+,$$

where  $\mathbb{E}$  denotes the expectation with respect to  $P_\pi(\cdot, t)$ , follows immediately. It is easy to verify that the above relation can be concisely written as an LMI involving the moment trajectory of  $P_\pi$ . Concretely, we can write

$$\mathbf{M}_f(\mathbf{y}(t)) \succeq \mathbf{0}, \tag{LMI}$$

where  $\mathbf{M}_f : \mathbb{R}^{n_L+n_H} \rightarrow \mathbb{S}^{\binom{n+d}{d}}$  is a linear map. The precise structure of  $\mathbf{M}_f$  depends on  $f$  and is immaterial for all arguments presented in this paper; however, the interested reader is referred to Lasserre [46] or Dowdy and Barton [21] for a detailed and formal description of the structure of  $\mathbf{M}_f$ . As clear from the above argument, the construction of valid LMIs of form (LMI) relies merely on polynomials that are non-negative on  $X$ . For stochastic chemical systems, natural choices of such polynomials, reflecting that  $P_\pi$  is non-negative and in particular not supported on states with negative molecular counts, are  $f(\mathbf{x}) = 1$  and  $f(\mathbf{x}) = x_i$  for  $i = 1, \dots, n$  [21, 25, 43, 59]. More generally, the support of  $P_\pi(\cdot, t)$  on any basic closed semialgebraic set can be reflected this way, most importantly including the special cases of polyhedra and bounded integer lattices. To account for this flexibility while simplifying notation, we will make use of the following definition and shorthand notation.

**Definition 2.3** Let  $f_0, \dots, f_{n_p}$  be polynomials that are non-negative on the reachable set  $X$ . The convex cone described by LMIs generated by these polynomials is denoted by  $C(X)$ , i.e.,  $C(X) = \{\mathbf{y} \in \mathbb{R}^{n_L+n_H} \mid \mathbf{M}_{f_i}(\mathbf{y}) \succeq \mathbf{0}, i = 0, \dots, n_p\}$ .

Lastly, we note that the validity of LMIs of form (LMI) carries over to the generalized moments that are generated by non-negative test functions. To see this, observe that the linearity of  $\mathbf{M}_f$  implies that

$$\mathbf{M}_f(\mathbf{z}(g; t)) = \int_0^t g(\tau) \mathbf{M}_f(\mathbf{y}(\tau)) d\tau$$

holds. Now assuming  $g$  is non-negative on  $\mathbb{R}_+$  and applying Jensen’s inequality to the extended convex indicator function of the positive semidefinite cone,  $\mathbb{1}_{\mathbb{S}_+^\infty}$ , yields that

$$0 \leq \mathbb{1}_{\mathbb{S}_+^\infty}(\mathbf{M}_f(\mathbf{z}(g; t))) \leq \int_0^t g(\tau) \mathbb{1}_{\mathbb{S}_+^\infty}(\mathbf{M}_f(\mathbf{y}(\tau))) d\tau = 0$$

and hence shows that  $\mathbf{M}_f(\mathbf{z}(g; t)) \succeq \mathbf{0}$  must hold for any  $t \geq 0$  in analogy to Eq. (LMI).



### 2.3.2 Affine Constraints

The moment dynamics (mCME) give rise to affine constraints that the moments and generalized moments must satisfy [11, 13, 37, 41, 47, 48, 61]. To see this, consider a test function  $g \in \mathcal{A}([0, t_f])$  and final time  $t_f \leq t_T$ . Integrating  $\int_0^{t_f} g(t) \frac{dy_L}{dt}(t) dt$  by parts yields the following set of affine equations:

$$\mathbf{K} (g(t_f)\mathbf{y}(t_f) - g(0)\mathbf{y}(0)) = \mathbf{A}\mathbf{z}(g; t_f) + \mathbf{K}\mathbf{z}(g'; t_f). \quad (2)$$

We wish to emphasize here that the above constraints are vacuous if  $\mathbf{z}(g; t)$  and  $\mathbf{z}(g'; t)$  are no further restricted. This observation motivates necessary restrictions on the test function  $g$  to generate “useful” generalized moments. Recalling the discussion in Sect. 2.3, one may be tempted to argue that  $g$  and  $g'$  shall be non-negative (or non-positive) on  $[0, t_f]$  so that the generated generalized moments satisfy LMIs of form (LMI). In fact, Dowdy and Barton [22] as well as Sakurai and Hori [61] demonstrate that this is indeed a reasonable strategy; they use exponential and monomial test functions, respectively. However, in principle a wider range of test functions can be used. We defer the discussion of this issue to Sect. 3.

## 3 Tighter Bounds

### 3.1 An Optimal Control Perspective

Some of the conservatism in the original method of Dowdy and Barton [22] stems from the fact that moments are only constrained in an *integral* or *weak* sense; i.e.,  $\mathbf{z}(g; t_f) = \int_0^{t_f} g(\tau)\mathbf{y}(\tau) d\tau$  is constrained as opposed to  $\mathbf{y}(t)$  for all  $t \in [0, t_f]$ . This is potentially a strong relaxation as in fact the entire trajectory must satisfy the necessary moment conditions. Moreover, by Assumption 3, the moment trajectories remain bounded at all times, which, taken together with the fact that they satisfy the ODE (mCME), implies that they are guaranteed to be smooth. Using these two additional pieces of information, we argue that the following continuous-time optimal control problem provides an elementary starting point for addressing the question of how to bound the moment trajectories associated with a stochastic chemical system evaluated at a given time point  $t_f$ :

$$\begin{aligned} \inf_{\mathbf{y} \in C^\infty(\mathbb{R}_+)} \quad & y_j(t_f) & (\text{OCP}) \\ \text{s.t.} \quad & \frac{d\mathbf{y}_L}{dt}(t) = \mathbf{A}_L\mathbf{y}_L(t) + \mathbf{A}_H\mathbf{y}_H(t), \quad \forall t \in \mathbb{R}_+, \\ & \mathbf{y}(0) = \mathbf{y}_0, \\ & \mathbf{y}(t) \in C(X), \quad \forall t \in \mathbb{R}_+. \end{aligned}$$

Here, the lower-order moments  $\mathbf{y}_L$  act as the state variables, while the higher-order moments  $\mathbf{y}_H$  can be viewed as control inputs. Although the infinite-dimensional nature

of Problem (OCP) leaves it with little immediate practical relevance, this representation is conceptually informative. It is not hard to verify that the method proposed by Dowdy and Barton [22] provides a systematic way to construct tractable SDP relaxations of (OCP). However, Dowdy and Barton’s method does in no way reflect the dependence of  $y(t_f)$  on past values of  $y(t)$  other than  $y(0)$  nor the fact that  $y$  is smooth or even continuous. As we will show in the following, ideas from the numerical analysis of ODEs allow us to reflect these features in the form of new necessary moment conditions constructed based on a discretization of the time domain of the problem. These conditions in turn yield tighter SDP relaxations of Problem (OCP) than those constructed by Dowdy and Barton’s method [22].

### 3.2 A New Hierarchy of Necessary Moment Conditions

In this section, we present the key contribution of this article—a new hierarchy of convex necessary moment conditions that reflect the temporal causality and regularity conditions inherent to the moment trajectories associated with the distribution described by the CME. To provide some intuition for these results, we will first discuss special cases of the proposed conditions which admit a clear interpretation. To derive these special cases, we specifically draw on two common ideas for the numerical analysis and solution of ODEs: temporal discretization and the analysis of the Taylor expansion of the moment trajectories [9, 16].

Recall that the moment trajectory  $y_j(\cdot)$  must be infinitely differentiable on  $\mathbb{R}_+$  as all moment trajectories remain bounded by Assumption 3 and obey the linear time-invariant dynamics (mCME). As a consequence, the Taylor polynomial

$$T_l(y_j; t_1, t_2) = \sum_{k=0}^l \frac{(t_2 - t_1)^k}{k!} \frac{d^k y_j}{dt^k}(t_1) \tag{3}$$

and remainder

$$R_l(y_j; t_1, t_2) = \frac{1}{l!} \int_{t_1}^{t_2} (t_2 - t)^l \frac{d^{l+1} y_j}{dt^{l+1}}(t) dt \tag{4}$$

are well-defined for any  $0 \leq t_1 \leq t_2 < +\infty$  and order  $l \geq 0$ . Now two things can be observed. First, higher-order time derivatives of  $y_j$  as in Eqs. (3) and (4) are simply linear combinations of high-order moments due to the linear dynamics (mCME). Second, the remainder term (4) then can be characterized as a linear combination of generalized moments for the test function  $g_l(t) = \mathbb{1}_{[t_1, t_2]}(t)(t_2 - t)^l$ . Thus, if  $|j|$  and  $l$  are sufficiently small,<sup>1</sup> then  $T_l(y_j; t_1, t_2)$  and  $R_l(y_j; t_1, t_2)$  depend linearly on  $y(t_1)$  and  $z(g_l; t_2)$ ; formally,

$$\begin{aligned} T_l(y_j; t_1, t_2) &= \mathbf{c}_{l,j}(t_1, t_2)^\top y(t_1), \\ R_l(y_j; t_1, t_2) &= \mathbf{d}_{l,j}(t_1, t_2)^\top z(g_l; t_2), \end{aligned} \tag{5}$$

<sup>1</sup> if  $|j|$  or  $l$  grows too large, the Taylor polynomial or remainder depends linearly on moments of higher order than  $m + q$

for an appropriate choice of the coefficient vectors. Overall, this observation suggests to employ conditions of the form

$$y_j(t_2) = \mathbf{T}_l(y_j; t_1, t_2) + \mathbf{R}_l(y_j; t_1, t_2)$$

at different time points along the trajectory as necessary moment conditions. These conditions achieve exactly what we set out to do: they establish a connection between  $y_j(t_2)$  and the moments at any past time point  $t_1$  using the smoothness properties of the moment trajectories  $\mathbf{y}(t)$ . In fact, for the same reason similar conditions are also used to derive and analyze numerical integration routines for ODEs such as Runge–Kutta or linear multistep methods [9, 16]. Further, it is straightforward to see that analogous conditions are readily obtained for any generalized moment generated by a sufficiently smooth test function. The above conditions hence appear to be a promising starting point. From a practical perspective, however, they merely suggest a particular choice of (local) polynomial test functions of the form  $g_l(t) = \mathbb{1}_{[t_1, t_2]}(t)(t_2 - t)^l$  as suggested by the integral form of the remainder in Eq. (4). This claim is formalized in the following proposition.

**Proposition 3.1** *Let  $0 \leq t_1 \leq t_2 < +\infty$  and  $n_l \leq \lfloor \frac{m}{q} \rfloor$ . Further, consider test functions of the form  $g_l(t) = \mathbb{1}_{[t_1, t_2]}(t)(t_2 - t)^l$ . If  $\mathbf{y}_{t_1}, \mathbf{y}_{t_2} \in \mathbb{R}^{n_L+n_H}$  and  $\mathbf{z}_{g_l, t_2} \in \mathbb{R}^{n_L+n_H}$  satisfy*

$$\mathbf{K} (g_l(t_2)\mathbf{y}_{t_2} - g_l(t_1)\mathbf{y}_{t_1}) = \mathbf{A}\mathbf{z}_{g_l, t_2} - l\mathbf{K}\mathbf{z}_{g_{l-1}, t_2} \quad (6)$$

for  $l = 0, \dots, n_l$ , then  $\mathbf{y}_{t_1}, \mathbf{y}_{t_2}$  and  $\mathbf{z}_{g_l, t_2}$  also satisfy

$$y_{j, t_2} = \mathbf{c}_{l, j}(t_1, t_2)^\top \mathbf{y}_{t_1} + \mathbf{d}_{l, j}(t_1, t_2)^\top \mathbf{z}_{g_l, t_2} \quad (7)$$

for  $l = 0, \dots, n_l$  and  $j$  such that  $|j| \leq m - lq$ , where  $\mathbf{c}_{l, j}$  and  $\mathbf{d}_{l, j}$  are defined as in Eq. (5).

**Proof** The proof is deferred to Appendix A.  $\square$

**Remark 3.1** Note that Condition (6) is analogous to Condition (2) as obtained for the test function  $g_l$  with a shifted origin, hence it is a necessary moment condition. Further, we wish to emphasize that Condition (6) is in general more stringent than Condition (7) as is made clear in the proof.

Beyond a specific choice of test functions, the above considerations motivate a broader strategy to generate necessary moment conditions that reflect causality. This strategy can be summarized as “discretize and constrain” and was previously shown to be effective in other contexts [17]. Instead of imposing Condition (2) on the entire time horizon  $[0, t_f]$  as proposed by Dowdy and Barton [22], the time horizon can be partitioned into  $n_\top$  subintervals  $[t_{i-1}, t_i]$  with  $0 = t_0 < t_1 < \dots < t_{n_\top} = t_f$  on which analogous conditions obtained from integrating  $\int_{t_{i-1}}^{t_i} g(\tau)\mathbf{K} \frac{d\mathbf{y}}{d\tau}(\tau) d\tau$  by parts can be imposed:

$$\begin{aligned} & \mathbf{K}(g(t_i)\mathbf{y}(t_i) - g(t_{i-1})\mathbf{y}(t_{i-1})) = \\ & \mathbf{A}(z(g; t_i) - z(g; t_{i-1})) + \mathbf{K}(z(g'; t_i) - z(g'; t_{i-1})). \end{aligned} \tag{8}$$

While by itself this does not provide any restriction over Condition (2), the following observation makes it worthwhile: the generalized moments generated by a non-negative test function  $g$  form a monotonically increasing sequence with respect to the convex cone  $C(X)$ . This follows immediately from the definition of  $z(g; t)$  and Jensen’s inequality as described in Sect. 2.3; formally,

$$z(g; t_i) - z(g; t_{i-1}) \in C(X), \quad i = 1, \dots, n_T \tag{9}$$

are necessary moment conditions. Conditions (8) & (9) are generally a non-trivial restriction of Condition (2) &  $z(g; t_f) \in C(X)$  as employed by Dowdy and Barton [22]. To see this, simply observe that we recover Eq. (2) by summing Eq. (8) over  $i = 1, \dots, n_T$  and likewise obtain

$$z(g; t_f) = \sum_{i=1}^{n_T} z(g; t_i) - z(g; t_{i-1}) \in C(X),$$

using that  $C(X)$  is a convex cone and  $z(g; 0) = \mathbf{0}$  by definition.

The above-described strategies lend themselves to generalization in terms of a hierarchy of necessary moment conditions. This generalization can be performed in several equivalent ways. Next we will present one such generalization utilizing a concept which we refer to as iterated generalized moments.

**Definition 3.1** Let  $z_j(g; t)$  be the  $j$ th generalized moment as per Definition 2.2. Then, the iterated generalized moment of Level  $l \geq 0$  is defined by

$$z_j^l(g; t) = \begin{cases} \int_0^t z_j^{l-1}(g; \tau) d\tau, & l \geq 1 \\ g(t)y_j(t), & l = 0 \end{cases}.$$

For the sake of simplified notation and analysis, it will further prove useful to introduce the left and right integral operators  $I_L, I_R : \mathcal{C}(\mathbb{R}^2) \rightarrow \mathcal{C}(\mathbb{R}^2)$  given by

$$(I_L f)(t_1, t_2) = \int_{t_1}^{t_2} f(t_1, t) dt \text{ and } (I_R f)(t_1, t_2) = \int_{t_1}^{t_2} f(t, t_2) dt.$$

For vector-valued functions,  $I_L$  and  $I_R$  shall be understood as being applied componentwise.

With these two notions in hand, a hierarchy of necessary moment conditions can be constructed by repeatedly integrating Eqs. (8) and (9) over an increment of the time domain partition. The resultant hierarchy of necessary moment conditions is summarized in the following proposition.

**Proposition 3.2** Let  $t_T > 0$  and consider a non-negative test function  $g \in \mathcal{AC}([0, t_T])$ . Further, let  $\mathbf{y}$  be the truncated sequence of moment trajectories associated with the solution of Eq. (CME), and  $\mathbf{z}^l$  be the corresponding iterated generalized moments. Then, the following conditions hold for any  $l \geq 1$ :

(i) For any  $t \in [0, t_T]$  it holds that

$$\mathbf{A}\mathbf{z}^l(g; t) + \mathbf{K}\mathbf{z}^l(g'; t) = \mathbf{K} \left( \mathbf{z}^{l-1}(g; t) - \frac{t^{l-1}}{(l-1)!} g(0)\mathbf{y}(0) \right).$$

(ii) Let  $\mathbf{f}(x, y) = \mathbf{z}^1(g; y) - \mathbf{z}^1(g; x)$ . Then, for any  $0 \leq t_1 \leq t_2 \leq t_T$  and  $k \in \{0, \dots, l-1\}$  it holds that  $(I_L^{l-1-k} I_R^k \mathbf{f})(t_1, t_2) \in C(X)$ .

**Proof** It is easily verified that Condition (i) is obtained from integrating Eq. (2)  $l-1$  times. Validity of Condition (ii) follows by a similar inductive argument: Since  $\mathbf{y}(t) \in C(X)$  for all  $t \in [0, t_T]$ , it follows by non-negativity of  $g$  on  $[0, t_T]$  and Jensen's inequality that

$$\mathbf{z}^1(g; t_2) - \mathbf{z}^1(g; t_1) = \int_{t_1}^{t_2} g(t)\mathbf{y}(t) dt \in C(X)$$

for any  $0 \leq t_1 \leq t_2 \leq t_T$ . Now suppose Condition (ii) is satisfied for  $l-1$ . Then, it follows by Jensen's inequality that for any  $0 \leq t_1 \leq t_2 \leq t_T$  and  $k = 0, \dots, l-2$

$$(I_L^{l-1-k} I_R^k \mathbf{f})(t_1, t_2) = \int_{t_1}^{t_2} (I_L^{l-2-k} I_R^k \mathbf{f})(t_1, t) dt \in C(X).$$

For  $k = l-1$ , an analogous argument applies. □

Before we proceed, a few remarks are in order to contextualize this result.

**Remark 3.2** Choosing  $l = 1$ ,  $t_1 = 0$ ,  $t_2 = t_f$  and exponential test functions of the form  $g(t) = e^{\rho(t_T-t)}$  reproduces the necessary moment conditions proposed by Dowdy and Barton [22].

**Remark 3.3** Regarding Condition (ii), one might be tempted to argue that any permutation of the operator products  $I_L$  and  $I_R$  of length  $l-1$  applied to  $\mathbf{f}(x, y) = \mathbf{z}^1(g; y) - \mathbf{z}^1(g; x)$  gives rise to a new valid necessary moment condition. It can be confirmed, however, that  $I_L$  and  $I_R$  commute such that Condition (ii) is invariant under permutation of  $I_L$  and  $I_R$  (see Appendix B).

**Remark 3.4** We wish to emphasize that Conditions (i) and (ii) depend affinely on the iterated generalized moments up to Level  $l$  evaluated at  $t_1$  and  $t_2$ , respectively. Accordingly, they preserve the computational advantages of Dowdy and Barton's necessary moment conditions. To avoid notational clutter in the remainder of this article we will disguise this fact and concisely denote the left-hand side of Condition (ii) by

$$\Omega_{l,k} \left( \left\{ z^i(g; t_1) \right\}_{i=1}^l, \left\{ z^i(g; t_2) \right\}_{i=1}^l, t_1, t_2 \right).$$

An explicit algebraic expression for  $\Omega_{l,k}$  is provided in Appendix D.

**Remark 3.5** For the  $\mathbf{0}$ -th-generalized moments, additional constraints arise from the definition as

$$z^l_0(g; t) = \begin{cases} \int_0^t z_0^{l-1}(g; \tau) d\tau, & l \geq 1 \\ g(t), & l = 0 \end{cases}$$

can be evaluated explicitly.

It is crucial to emphasize that a careful choice of test functions is critical to endow the conditions put forth in Proposition 3.2 with restrictive power. In particular, Condition (i) in Proposition 3.1 is effectively unrestrictive unless  $z^l(g'; t)$  can be further constrained. Fortunately, one can draw from rich function classes including polynomials [61], trigonometric functions [11, 44] and exponentials [22] to assemble a set of test functions that render Condition (i) in Proposition 3.1 “self-contained.” For such test function sets,  $z^l(g'; t)$  reduces to linear combinations of the generalized moments generated by the test function set itself. The following examples showcase how such test function sets can be constructed in practice from exponential and monomial test functions.

**Example 3.1** Consider test functions of the form  $g(t) = e^{\rho t}$ . Then, the associated generalized moments satisfy  $z^l(g'; t) = \rho z^l(g; t)$  due to the linearity of  $z^l$  in its first argument. Condition (i) of Proposition 3.2 thus reduces to

$$(\mathbf{A} + \rho \mathbf{K}) z^l(g; t) = \mathbf{K} \left( z^{l-1}(g; t) - \frac{t^{l-1}}{(l-1)!} g(0) \mathbf{y}(0) \right).$$

**Example 3.2** Consider a set of test functions of the form  $g_k(t) = t^k$  for  $k = 0, \dots, n$ . By linearity of  $z^l$  in its first argument, the identity  $z^l(g'_k; t) = k z^l(g_{k-1}; t)$  must hold. Accordingly, Condition (i) in Proposition 3.2 reduces to

$$\mathbf{A} z^l(g_k; t) + k \mathbf{K} z^l(g_{k-1}; t) = \mathbf{K} \left( z^{l-1}(g_k; t) - \frac{t^{l-1}}{(l-1)!} g_k(0) \mathbf{y}(0) \right).$$

The preceding discussion and examples indicate that the span of an appropriate test function set should be closed under differentiation. The following proposition formalizes this guideline.

**Proposition 3.3** *Let  $F$  be a finite set of test functions such that  $\text{span}(F)$  is closed under differentiation. Then, for any  $f \in F$ , there exists a linear map  $\Gamma_f$  such that Condition (i) in Proposition 3.2 is equivalent to*

$$\Gamma_f \left( \left\{ z^l(g; t) \right\}_{g \in F} \right) = \mathbf{K} \left( z^{l-1}(f; t) - \frac{t^{l-1}}{(l-1)!} f(0) \mathbf{y}(0) \right).$$

We omit the elementary proof of Proposition 3.3 and instead refer back to Examples 3.1 and 3.2 for how the linear map  $\Gamma_f$  can be constructed in practice.

Another issue pertaining to the choice of test functions is the requirement of non-negativity in Proposition 3.2. This problem can be alleviated by a simple reformulation and shift of the time horizon in Proposition 3.2. For example, if a test function  $g$  is non-negative on  $[0, t_+]$  and non-positive on  $[t_+, t_T]$ , we can simply consider the two test functions  $g_+(t) = \mathbb{1}_{[0, t_+]}(t)g(t)$  and  $g_-(t) = -\mathbb{1}_{[t_+, t_T]}(t)g(t)$  in place of  $g$  and impose the necessary moment conditions on the intervals  $[0, t_+]$  and  $[t_+, t_T]$ , respectively. This construction naturally extends to test functions with any finite number of sign changes.

We conclude this section by establishing some compelling properties of the hierarchy of necessary moment conditions put forward in Proposition 3.2. On the one hand, Conditions (i) and (ii) in Proposition 3.2 include the conditions considered in Proposition 3.1 as special cases. So in particular, they enforce consistency with higher-order Taylor expansions of the true moment trajectories as discussed in the beginning of this section. The following corollary to Proposition 3.2 formalizes this claim.

**Corollary 3.1** *Let  $n_I \in \mathbb{Z}_+$  and  $t_T > 0$  be fixed. Further, suppose  $g \in \mathcal{AC}([0, t_T])$  is non-negative, and let  $\mathbf{y}$  and  $\mathbf{z}^l(\cdot; \cdot)$  be arbitrary functions such that  $\mathbf{z}^l(\cdot; \cdot)$  is linear in the first argument and  $\mathbf{z}^0(g; t) = g(t)\mathbf{y}(t)$  holds. Fix  $0 \leq t_1 \leq t_2 \leq t_T$  and define  $h_l(t) = \mathbb{1}_{[t_1, t_2]}(t)(t_2 - t)^l$  for  $l = 0, 1, \dots, n_I$ . If Conditions (i) and (ii) of Proposition 3.2 are satisfied by  $\{\mathbf{z}^l(g; t_i)\}_{l=0}^{n_I+1}$  for  $i = 1, 2$ , then there exist functions  $\mathbf{z}(\cdot; \cdot)$  that are linear in the first argument, and satisfy*

$$\mathbf{K}(h_l(t_2)g(t_2)\mathbf{y}(t_2) - h_l(t_1)g(t_1)\mathbf{y}(t_1)) = \mathbf{A}\mathbf{z}(h_l g; t_2) + \mathbf{K}\mathbf{z}((h_l g)'; t_2) \quad (10)$$

and

$$\mathbf{z}(h_l g; t_2) \in C(X) \quad (11)$$

for all  $l \in \{0, \dots, n_I\}$ .

**Proof** The proof is deferred to Appendix C. □

**Remark 3.6** To see the connection between Condition (10) and Condition (6) in Proposition 3.1, simply consider the case where  $g(t) = 1$ . Moreover, note that Corollary 3.1 also shows that necessary moment conditions of the form of (8) & (9) are implied as they are recovered for  $l = 0$  since we can simply identify  $\mathbf{z}(h_0 g; t_2)$  with  $\mathbf{z}^1(g; t_2) - \mathbf{z}^1(g; t_1)$ .

On the other hand, the proposed necessary moment conditions display benign scaling behavior. Condition (ii) in Proposition 3.2 scales quadratically with respect to the level  $l$  of the hierarchy, independent of the state space dimension. Moreover, a strengthening of the conditions is obtained by refining the partition of the time horizon. This strengthening mechanism does not only provide a desirable degree of flexibility but also enjoys linear scaling with respect to the number of intervals in the partition which of course is independent of the state space dimension. In particular, Condition (ii) must

only be imposed between the endpoints of each interval of the partition to ensure it holds between any two endpoints. This claim is formalized in the following Corollary.

**Corollary 3.2** *Let  $0 \leq t_1 \leq t_2 \leq t_3 < +\infty$  and  $n_I$  be a fixed positive integer. Suppose  $\{z^s\}_{s=1}^{n_I}$  is a set of functions such that*

$$\Omega_{l,k} \left( \{z^s(t_i)\}_{s=1}^l, \{z^s(t_{i+1})\}_{s=1}^l, t_i, t_{i+1} \right) \in C(X)$$

for all  $i \in \{1, 2\}$  and  $k, l \in \mathbb{Z}_+$  such that  $k < l \leq n_I$ . Then,

$$\Omega_{l,k} \left( \{z^s(t_1)\}_{s=1}^l, \{z^s(t_3)\}_{s=1}^l, t_1, t_3 \right) \in C(X)$$

holds for all  $k, l \in \mathbb{Z}_+$  such that  $k < l \leq n_I$ .

**Proof** The proof is deferred to Appendix D. □

### 3.3 An Augmented Semidefinite Program

In this section, we construct an SDP based on the hierarchy of necessary moment conditions put forth in Proposition 3.2. The optimal value of this SDP furnishes bounds on the moment solutions of Eq. (CME) at a given time point  $t_f \in [0, t_T]$ . To that end, we consider the moment truncation order  $m$  to be fixed and the following user choices as known:

- (i)  $T = \{t_1, \dots, t_{n_T}\}$ —A finite, ordered set of time points framing the partition of the time horizon. For notational simplicity, we assume that  $0 < t_1 < t_2 < \dots < t_{n_T} \leq t_T$  and  $t_f \in T$ .
- (ii)  $F = \{g_1, \dots, g_{n_F}\}$ —A finite set of test functions that satisfies the hypotheses of Propositions 3.2 and 3.3.
- (iii)  $n_I$ —A non-negative integer controlling the hierarchy level in Proposition 3.2.

These quantities parametrize a spectrahedron  $S(F, T, n_I)$  described by the necessary moment conditions of Proposition 3.2 as imposed for all test functions in  $F$ , at all time points in  $T$  and for all hierarchy levels up to  $n_I$ . In the formulation of  $S(F, T, n_I)$ , however, we use a slightly different but equivalent formulation of Condition (i) of Proposition 3.2. The reason for this modification is that it results in weakly coupled conditions that allow the resultant SDPs to be decomposed in line with the temporally causal structure of the constraints as we will discuss in Sect. 4.1. Details on this reformulation can be found in Appendix E.  $S(F, T, n_I)$  is explicitly stated below; for the sake of concise notation we introduced the shorthand  $n(t)$  for the left adjacent time point of  $t \in T$ , i.e.,  $n(t_i) = t_{i-1}$  for  $i = 2, \dots, n_T$  and  $n(t_1) = 0$ .



$$S(F, T, n_I) = \left\{ \left. \begin{array}{l} z_{g,t}^0 = g(t)y_t, \quad \forall (g, t) \in F \times T, \\ y_t \in C(X), \quad \forall t \in T, \\ \Gamma_g(\{z_{f,t}^l\}_{f \in F}) = \begin{cases} \mathbf{K} \left( z_{g,t}^{l-1} - \frac{t^{l-1}}{(l-1)!} g(0)y_0 \right), & \text{if } t = t_1 \\ \left(\frac{t}{n(t)}\right)^{l-1} \Gamma_g(\{z_{f,n(t)}^l\}_{f \in F}) \\ + \mathbf{K} \left( z_{g,t}^{l-1} - \left(\frac{t}{n(t)}\right)^{l-1} z_{g,n(t)}^{l-1} \right), & \text{if } t \neq t_1 \end{cases} \\ \forall (g, t, l) \in F \times T \times \{1, \dots, n_I\}, \\ \Omega_{l,k}(\{z_{g,n(t)}^s\}_{s=1}^l, \{z_{g,t}^s\}_{s=1}^l, n(t), t) \in C(X), \\ \forall (g, t) \in F \times T \text{ and } \forall k, l \in \mathbb{Z}_+ \text{ such that } k < l \leq n_I \end{array} \right\}.$$

By construction, the set  $S(F, T, n_I)$  contains the sequences  $\{y(t) : t \in T\}$  and  $\{z^l(g; t) : (g, t, l) \in F \times T \times \{0, \dots, n_I\}\}$  as generated by the true moment trajectories associated with the solution of Eq. (CME). Another piece of information that can be used to further restrict the set of candidates for the true moment solutions to Eq. (CME) is information about the moments of the initial distribution. We know for example from the definition that any iterated generalized moment  $z^l(g; t)$  for  $l \geq 1$  must vanish at  $t = 0$ . Moreover, one usually has specific information about the initial distribution of the system state, hence also about  $y(0)$ . Here, we assume that the initial moments and iterated generalized moments are confined to a spectrahedral set denoted by  $S_0(F, n_I)$ . In the common setting in which the moments of the initial distribution are known exactly,  $S_0(F, n_I)$  would be given by

$$S_0(F, n_I) = \left\{ y_0, \{z_{g,0}^l\} \mid \begin{array}{l} y_0 = y(0), \\ z_{g,0}^l = \mathbf{0}, \quad \forall (g, l) \in F \times \{1, \dots, n_I\} \end{array} \right\}.$$

Albeit adding the corresponding constraints to the description of  $S(T, F, n_I)$  appears natural, we deliberately choose to reflect this piece of information separately. Our motivation for this distinction is twofold: On the one hand, we want to emphasize that the presented approach naturally extends to the setting of uncertain or imperfect knowledge of the moments of the initial distribution of the system. Specifically, if the moments of the initial distribution are not known exactly, however, known to be confined to a spectrahedral set, the proposed bounding procedure applies without modification. On the other hand, we will argue in Sect. 4.1 that the arising optimization problems lend themselves to be decomposed according to the temporal structure of the constraints. The distinction in notation made here will simplify our exposition there.

The following theorems finally summarize the key feature of our proposed bounding approach—the ability to generate a sequence of monotonically improving, practically computable bounds on the moment trajectories associated with the solution of Eq. (CME). Theorem 3.1 shows that these bounds can be obtained by way of solving an SDP.

**Theorem 3.1** *Let  $y$  denote the transient moments as described in Definition 2.1. Further, for any time point  $t_f \in T$  and multi-index  $|i| \leq m$ , define*

$$\begin{aligned}
 y_{i,t_f}^* &= \inf_{\{\mathbf{y}_t\}, \{\mathbf{z}_{g,t}^l\}} y_{i,t_f} && \text{(SDP)} \\
 \text{s.t.} & \quad (\{\mathbf{y}_t\}, \{\mathbf{z}_{g,t}^l\}) \in S(\mathbb{F}, \mathbb{T}, n_I), \\
 & \quad (\mathbf{y}_0, \{\mathbf{z}_{g,0}^l\}) \in S_0(\mathbb{F}, n_I).
 \end{aligned}$$

Then,  $y_{i,t_f}^* \leq y_i(t_f)$ .

**Proof** Set  $\mathbf{y}_t = \mathbf{y}(t)$  for all  $t \in \mathbb{T}$  and  $\mathbf{z}_{g,t}^l = \mathbf{z}^l(g; t)$  for all  $(g, t, l) \in \mathbb{F} \times \mathbb{T} \times \{0, \dots, n_I\}$  with  $\mathbf{z}^l(g; t)$  as in Definition 3.1. By Proposition 3.2  $(\{\mathbf{y}_t\}, \{\mathbf{z}_{g,t}^l\}) \in S(\mathbb{F}, \mathbb{T}, n_I)$ , and the result follows.  $\square$

**Remark 3.7** For equal truncation orders and choice of  $C(X)$ , Remark 3.2 implies that the bounds obtained from (SDP) are at least as tight as those obtained by the approach of Dowdy and Barton [22].

**Remark 3.8** The lower bound  $y_{i,t_f}^*$  can be evaluated using off-the-shelf solvers for SDPs such as MOSEK [6], SeDuMi [68] or SDPT3 [69].

**Remark 3.9** Similar problems as (SDP) can be formulated to bound properties that can be described in terms of moments of non-negative measures on the reachable set; examples include variances [21], the volume of a confidence ellipsoids [60] and the mass that the probability measure assigns to a semialgebraic set [21].

The formulation of (SDP) provides several mechanisms to improve the bounds by adjusting the parameters  $\mathbb{T}$ ,  $\mathbb{F}$  and  $n_I$ . Theorem 3.2 shows that appropriate adjustments lead to a sequence of monotonically improving bounds.

**Theorem 3.2** Let  $y_{i,t_f}^*$  be defined as in Theorem 3.1. Let  $\tilde{t} \in [0, t_T]$  and define  $\tilde{\mathbb{T}} = \mathbb{T} \cup \{\tilde{t}\}$ . Further, let  $\tilde{g}$  be an absolutely continuous function that is non-negative on  $[0, t_T]$  and define  $\tilde{\mathbb{F}} = \mathbb{F} \cup \{\tilde{g}\}$ . Then,

$$\begin{aligned}
 y_{i,t_f}^* &\leq \inf_{\{\mathbf{y}_t\}, \{\mathbf{z}_{g,t}^l\}} y_{i,t_f} \\
 \text{s.t.} & \quad (\{\mathbf{y}_t\}, \{\mathbf{z}_{g,t}^l\}) \in S(\tilde{\mathbb{F}}, \tilde{\mathbb{T}}, n_I), \\
 & \quad (\mathbf{y}_0, \{\mathbf{z}_{g,0}^l\}) \in S_0(\tilde{\mathbb{F}}, n_I).
 \end{aligned} \tag{12}$$

Likewise,

$$\begin{aligned}
 y_{i,t_f}^* &\leq \inf_{\{\mathbf{y}_t\}, \{\mathbf{z}_{g,t}^l\}} y_{i,t_f} \\
 \text{s.t.} & \quad (\{\mathbf{y}_t\}, \{\mathbf{z}_{g,t}^l\}) \in S(\mathbb{F}, \mathbb{T}, n_I + 1), \\
 & \quad (\mathbf{y}_0, \{\mathbf{z}_{g,0}^l\}) \in S_0(\mathbb{F}, n_I + 1).
 \end{aligned} \tag{13}$$

**Proof** Inequality (12) is clearly true if  $\tilde{t} \in \mathbb{T}$  and  $\tilde{g} \in \mathbb{F}$ . If  $\tilde{t} \notin \mathbb{T}$  and/or  $\tilde{g} \notin \mathbb{F}$ , any feasible point of the right-hand side of (12) can be used to construct a feasible point

**Table 1** Scaling of (SDP)

|       | #variables                       | #affine constraints              | #LMI       | LMI size   |
|-------|----------------------------------|----------------------------------|------------|--|
| F     | $O( F )$                         | $O( F )$                         | $O( F )$   | $O(1)$   |
| T     | $O( T )$                         | $O( T )$                         | $O( T )$   | $O(1)$   |
| $n_I$ | $O(n_I)$                         | $O(n_I)$                         | $O(n_I^2)$ | $O(1)$   |
| $m$   | $O\left(\binom{m+q+n}{n}\right)$ | $O\left(\binom{m+q+n}{n}\right)$ | $O(1)$     | $O\left(\binom{\lfloor(m+q)/2\rfloor+n}{n}\right)$ |

of (SDP); simply remove the decision variables that correspond to time point  $\tilde{t}$  and/or test function  $\tilde{g}$ . Similarly, removing the iterated generalized moments of Level  $n_I + 1$  of the right-hand side of (13) yields a feasible point for (SDP).  $\square$

**Remark 3.10** Increasing the truncation order also gives rise to monotonically improving bounds. For the sake of brevity, we omit a formal statement and proof here as many easily adapted results of this type exist; see for example Corollary 6 in Kuntz et al. [43].

Theorems 3.1 and 3.2 establish that the proposed necessary moment conditions provide a practical way to compute hard bounds on the transient moments associated with stochastic chemical systems alongside several mechanisms to tighten the bounds. A natural question in this context is whether these bounds converge to the true transient moments as the bound tightening mechanisms are taken to the limit. While we find that the bounds often become tight enough to be of practical value, we cannot claim convergence. This contrasts with many applications of the standard moment-sum-of-squares hierarchy, such as polynomial optimization [45], solving nonlinear PDEs [41], option pricing [48] and deterministic optimal control [37, 47], where convergence guarantees can be established under regularity conditions. The main obstacle obstructing the extension of these results to bounding schemes for stochastic chemical systems is the discrete nature of the state space of such systems. On the one hand, the state space of open systems is an unbounded integer lattice and hence not basic closed semialgebraic. This violates a key assumption underpinning existing convergence guarantees. On the other hand, even though the state space of closed systems is a bounded integer lattice and thus also a basic closed semialgebraic set, it typically features an exceedingly large number of elements, rendering the reflection of all individual elements of this set through LMIs intractable. To the best of our knowledge, the only moment bounding scheme for stochastic chemical systems that defies these complications and provides convergence guarantees is that of Kuntz et al. [43]. However, their construction relies on enumerating the discrete states through growing, finite truncations of the state space, complemented by moment bounds accounting for the truncated states. As such, their bounding scheme forfeits the perks of a purely moment-based description and is therefore more limited in practice.

We conclude this section with a brief discussion of the scalability of (SDP). Table 1 summarizes how the number of variables, affine constraints and LMIs as well as their dimension scales with F, T,  $n_I$  and the truncation order  $m$ . The results demonstrate the value of the proposed formulation if the number of species  $n$  in the system

under investigation is large. In that case, the bound tightening mechanisms offered by adjusting  $F$ ,  $T$  and  $n_I$  scale much more moderately than increasing the truncation order. Furthermore, it should be emphasized that the invariance of LMI size with respect to  $F$ ,  $T$  and  $n_I$  is a particularly desirable property to achieve scalability of SDP hierarchies in practice [1, 3]. Lastly, it is worth noting that moment-based SDPs are notorious for becoming numerically ill-conditioned as the truncation order increases. Thus, the presented hierarchy provides a mechanism to circumvent this issue to an extent.

## 4 Practical Considerations

### 4.1 Leveraging Causality for Decomposition

Techniques for the efficient numerical integration of ODEs hinge fundamentally on the causality that is inherent to the solution of ODEs. It enables the original problem, namely integration over a long time horizon, to be decomposed into a sequence of simpler, more tractable subproblems, each corresponding to integration over only a small fraction of the time horizon. In this section, we discuss how the structure of the presented optimization problems can be exploited in a similar spirit. Additionally, we show that such exploitation of structure gives rise to a mechanism for trading off tractability and bound quality.

Suppose we are interested in computing moment bounds at the end of a *long* time horizon  $[0, t_f]$ . In light of the arguments made in Sect. 3.2, it is reasonable to expect that the set  $T$  should ideally be populated with a large number of time points in this setting. Accordingly, solving the resultant optimization problem in one go may become prohibitively costly, even despite the benign scaling of the SDP size with respect to  $|T|$ . As alluded to in the beginning of this section, this limitation may be circumvented by decomposing the problem into a sequence of simpler subproblems each of which cover only a fraction of the time horizon. To that end, suppose that  $T = \{t_1, \dots, t_{n_T}\}$  is ordered with  $t_{n_T} = t_f$ , and let  $t_0 = 0$ . Further consider the subsets  $T_1, \dots, T_{n_T}$  of  $T$  such that  $T_k = \{t_k\}$ . We now define

$$S_k = \begin{cases} S_0(F, n_I), & \text{if } k = 0, \\ \left\{ \left. \begin{array}{l} \mathbf{y}_{t_k}, \{\mathbf{z}_{g,t_k}^l\} \\ \exists(\mathbf{y}_{t_{k-1}}, \{\mathbf{z}_{g,t_{k-1}}^l\}) \in S_{k-1} \text{ such that} \\ (\mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}, \{\mathbf{z}_{g,t_{k-1}}^l, \mathbf{z}_{g,t_k}^l\}) \in S(F, T_k, n_I) \end{array} \right\}, & \text{if } k \geq 1. \end{cases}$$

At this point, it is worth emphasizing the meaning of each  $S_k$  and how its construction directly exploits the way we formulated the necessary moment conditions in  $S(F, T, n_I)$ . To that end, note that each condition in  $S(F, T, n_I)$  only links variables corresponding to adjacent time points. As a consequence, the set  $S(F, T_k, n_I)$  constrains only the variables  $(\{\mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}\}, \{\mathbf{z}_{g,t_{k-1}}^l, \mathbf{z}_{g,t_k}^l\})$ . By construction of  $S_k$ , we project out the variables  $(\mathbf{y}_{t_{k-1}}, \{\mathbf{z}_{g,t_{k-1}}^l\})$  while imposing their membership in  $S_{k-1}$ . It follows by induction that  $S_k$  precisely describes the projection of  $S(F, \cup_{i=1}^k T_i, n_I)$  onto the variables  $(\mathbf{y}_{t_k}, \{\mathbf{z}_{g,t_k}^l\})$  under the condition that  $(\mathbf{y}_0, \{\mathbf{z}_{g,0}\}) \in S_0(F, n_I)$ . By this argument, it follows that the original problem (SDP) is equivalent to the following reduced space formulation:

$$\begin{aligned} & \inf_{\mathbf{y}_{t_f}, \{\mathbf{z}_{g,t_f}^l\}} \quad y_{i,t_f} \\ & \text{s.t.} \quad (\mathbf{y}_{t_f}, \{\mathbf{z}_{g,t_f}^l\}) \in S_{n_T}, \end{aligned}$$

where all decision variables that correspond to time points before  $t_f$  have been projected out. It should be clear that the above optimization problem provides a computational advantage over the original problem only if the set  $S_{n_T}$  can be represented, or at least tightly approximated, in a “simple” way. To that end, we suggest to successively compute conic outer approximations of the projections  $S_k$  according to Algorithm 1.

---

### Algorithm 1 Successive Overapproximation

---

- 1: **procedure**  $\text{SO}(S_0, S_1, \dots, S_{n_T})$
- 2:   Set  $\tilde{S}_0 = S_0$
- 3:   **for**  $k = 1, \dots, n_T$  **do**
- 4:     Compute conic overapproximation

$$\tilde{S}_k \supset \left\{ \mathbf{y}_{t_k}, \{\mathbf{z}_{g,t_k}^l\} \mid \begin{array}{l} \exists (\mathbf{y}_{t_{k-1}}, \{\mathbf{z}_{g,t_{k-1}}^l\}) \in \tilde{S}_{k-1} \text{ such that} \\ \left( \begin{array}{l} \{\mathbf{y}_{t_{k-1}}, \mathbf{y}_{t_k}\} \\ \{\mathbf{z}_{g,t_{k-1}}^l, \mathbf{z}_{g,t_k}^l\} \end{array} \right) \in S_k \end{array} \right\}$$

- 5:   **end for**
  - 6:   **return**  $\tilde{S}_{n_T}$
  - 7: **end procedure**
- 

Note that Algorithm 1 parallels the decomposition approach taken in classical numerical integration of ODEs: the task of finding moment bounds at the end of the time horizon  $[0, t_f]$  is decomposed into a sequence of smaller subproblems corresponding to finding moment bounds over smaller subintervals of the horizon, and each subproblem requires the solution of the previous subproblem as input data. In other words, Algorithm 1 propagates the moment bounds forward in time, successively subinterval by subinterval, in the same way as a numerical integrator propagates values of the state of a dynamical system forward in time.

We conclude this section with some final remarks. First, we would like to emphasize that the specific choices of the subdomains  $T_k$  made in this section are made purely for clarity of exposition. The partition can be chosen as coarse as desired, i.e., each  $T_k$  can comprise multiple time points, only requiring minimal adjustments of Algorithm 1. Second, computing and representing the conic overapproximations in Algorithm 1 may be expensive, in particular if many moments are considered. For example, computing a polyhedral outer approximation of the positive semidefinite cone is known to converge exponentially slowly in the worst-case [15]. Second-order cone approximations perform better empirically [1, 4] and theoretically [12], however, are more expensive to compute and represent. On the other hand, it may not be necessary to find overapproximations that are globally tight but only near the optimal solution of the original problem. Finally, with decisions on accuracy of the overapproximation and the

coarseness of the partition of  $T$  required in Algorithm 1, one is left with mechanisms to trade off accuracy and computational cost.

### 4.2 Quantifying Approximation Quality

A natural question that arises from the formulation of Problem (SDP) is how to choose the parameters required for its construction, i.e., the sets  $F$  and  $T$ , and the level  $n_I$  of the proposed constraint hierarchy. We will show that an approximation of Problem (OCP) can provide useful guidance for these choices. Specifically, we show that an approximation of (OCP) provides rigorous information on the best attainable bounds given the truncation order  $m$  is fixed. To that end, recall that (OCP) requires optimization over an infinite-dimensional vector space, namely  $C^\infty(\mathbb{R}_+)$ . To overcome this challenge, we will make two restrictions. On the one hand, we will restrict our considerations to a compact interval  $[0, t_T]$  and, on the other hand, we will restrict the search space to the set of univariate polynomials up to a fixed but arbitrary maximum degree  $d \in \mathbb{Z}_+$ . Note that the latter restriction is in some sense arbitrarily weak as  $\mathbb{R}[t]$  is dense in  $C^\infty([0, t_T])$  [58].

The above-discussed restrictions enable the construction of a tractable approximation of (OCP) using the following result which can be traced back to the work of Nesterov [53] as well as Powers and Reznick [56].

**Proposition 4.1** *Let  $m, d$  be positive integers. If  $d$  is odd, let  $r = k = \frac{d+1}{2}m$ . Otherwise, let  $k = (\frac{d}{2} + 1)m$ , and  $r = \frac{dm}{2}$ . Then, there exist two linear maps  $\alpha : \mathbb{S}^k \rightarrow \mathbb{S}^m[t]$  and  $\beta : \mathbb{S}^r \rightarrow \mathbb{S}^m[t]$  such that the matrix polynomial  $X \in \mathbb{S}^m[t]$  satisfies  $X(t) \succeq \mathbf{0}$  on  $[0, 1]$  if and only if there exist two matrices  $Q_\alpha \in \mathbb{S}_+^k$  and  $Q_\beta \in \mathbb{S}_+^r$  such that  $X = \alpha(Q_\alpha) + \beta(Q_\beta)$ .*

The maps  $\alpha$  and  $\beta$  in Proposition 4.1 are remarkably simple and freely available software tools for sum-of-squares programming allow for simple, concise implementation. The interested reader is referred to Ahmadi and El Khadir [2, Proposition 2] for an explicit description of  $\alpha$  and  $\beta$  alongside a simple proof of Proposition 4.1.

Proposition 4.1 allows to construct a tractable restriction of (OCP) on a compact horizon. The following theorem which may be regarded as a special case of the results of Ahmadi and El Khadir [2] formalizes this claim.

**Theorem 4.1** *Let  $d \in \mathbb{Z}_+$ . Then, the following semi-infinite optimization problem*

$$\begin{aligned}
 & \inf_{y \in \mathbb{R}_d^{n_L+n_H}[t]} y_j(t_f) && \text{(pOCP)} \\
 \text{s.t.} & \quad \frac{dy_L}{dt}(t) = A_L y_L(t) + A_H y_H(t), \quad \forall t \in [0, t_T], \\
 & \quad y(0) = y_0, \\
 & \quad y(t) \in C(X), \quad \forall t \in [0, t_T].
 \end{aligned}$$

*is equivalent to a finite SDP.*

**Proof** First, note that all equality constraints in the above optimization problem require equality of polynomials of fixed maximum degree. Accordingly, equality can be enforced by matching the coefficients of the polynomials when expressed in a common basis which in turn can be done via finitely many affine equality constraints. Additionally, recall that  $C(X)$  is described in terms of finitely many LMIs. Thus, the constraint  $\mathbf{y}(t) \in C(X), \forall t \in [0, t_T]$  is as well by Proposition 4.1.  $\square$

Unfortunately, (pOCP) may be a strong restriction and often even infeasible. However, the formulation of (pOCP) can be further relaxed without giving up too much relevant information. Specifically, we propose to restrict the solution space to piecewise polynomial functions in analogy to the collocation approach to optimal control [18]. The following corollary to Theorem 4.1 formalizes this approach.

**Corollary 4.1** *Let  $d, n_T \in \mathbb{Z}_+$  and consider  $n_T + 1$  time points  $t_0, \dots, t_{n_T}$  such that  $0 = t_0 < t_1 < \dots < t_{n_T} \leq t_T$ . Further suppose that  $t_f \in [t_k, t_{k-1}]$  for some  $k$ . Then, the following semi-infinite optimization problem*

$$\begin{aligned} \inf_{\mathbf{y}^i \in \mathbb{R}_d^{n_L+n_H} [t]} \quad & \mathbf{y}_j^k(t_f) && \text{(pwpOCP)} \\ \text{s.t.} \quad & \frac{d\mathbf{y}_L^i}{dt}(t) = \mathbf{A}_L \mathbf{y}_L^i(t) + \mathbf{A}_H \mathbf{y}_H^i(t), \quad \forall t \in [t_{i-1}, t_i], \quad \forall i \in \{1, \dots, n_T\}, \\ & \mathbf{y}^i(t_i) = \mathbf{y}^{i+1}(t_i), \quad \forall i \in \{1, \dots, n_T - 1\}, \\ & \mathbf{y}^1(0) = \mathbf{y}_0, \\ & \mathbf{y}^i(t) \in C(X), \quad \forall t \in [t_{i-1}, t_i], \quad \forall i \in \{1, \dots, n_T\} \end{aligned}$$

is equivalent to a finite SDP. Further, (pwpOCP) is a valid restriction of (SDP).

**Proof** That (pwpOCP) is equivalent to a finite SDP follows immediately from Theorem 4.1. Further, let  $\{\mathbf{y}^i\}$  be feasible for (pwpOCP) and consider the piecewise polynomial obtained by parsing the  $\mathbf{y}^i$  together like

$$\tilde{\mathbf{y}}(t) = \mathbf{y}^i(t), \quad \forall t \in (t_{i-1}, t_i] \text{ and } \forall i \in \{1, \dots, n_T\}.$$

By construction  $\tilde{\mathbf{y}}$  satisfies Eq. (mCME) and  $\tilde{\mathbf{y}}(t) \in C(X), \forall t \in [0, t_f]$ . Accordingly, the iterated generalized moments obtained from  $\tilde{\mathbf{y}}$  satisfy Conditions (i) and (ii) in Proposition 3.2. Thus, it is straightforward to generate a feasible point for (SDP) from  $\tilde{\mathbf{y}}$ .  $\square$

Since (pwpOCP) is fully independent of the choice of  $F, T$  and  $n_I$ , it provides a way to check rigorously the approximation quality of (SDP) against the baseline of (OCP). This can guide the user choice of the truncation order  $m$  and the parameters  $F, T$  and  $n_I$ . Specifically, the difference of optimal values of (pwpOCP) and (SDP) quantifies the potential for improvements by adding elements to  $F$  and  $T$  versus moving to a higher level in the proposed hierarchy.

## 5 Examples

In this section, we present several case studies that demonstrate the effectiveness of the proposed bounding hierarchy. We put special emphasis on showcasing that the proposed method enables the computation of substantially tighter bounds than can be obtained by the method of Dowdy and Barton [22]. Throughout, we use the subscripts *DB* to indicate any results obtained with Dowdy and Barton's method [22] and the subscript *HB* for those generated with the method presented in this paper.

### 5.1 Preliminaries

#### 5.1.1 Reaction Kinetics

All considered reaction networks are modeled via the CME and are assumed to obey mass action kinetics such that Assumptions 1 and 2 are naturally satisfied.

#### 5.1.2 State Space & LMIs

Following Dowdy and Barton [22], we reduce the state space of every reaction network explicitly to the minimum number of independent species by eliminating reaction invariants. Further, we employ the same set of LMI generating polynomials as suggested by Dowdy and Barton [22]. The resultant LMIs of the form (LMI) reflect non-negativity of the probability measure as well as its support on states with non-negative molecular counts.

#### 5.1.3 Hierarchy Parameters

Applying the proposed bounding scheme requires the user to specify a range of parameters, namely the truncation order  $m$ , the hierarchy level  $n_I$ , the test function set  $F$  and the set of time points  $T$  used to discretize the time horizon. While all these hierarchy parameters can in principle be chosen arbitrarily (assuming the test functions satisfy the hypotheses of Propositions 3.2 and 3.3) and independently, a careful choice is essential to achieve a good trade-off between bound quality and computational cost. At present we are not aware of a systematic way of choosing the hierarchy parameters optimally in that sense; however, our findings indicate that the following set of simple heuristics lead to good performance in practice.

- The set of time points  $T$  is chosen by equidistantly discretizing the entire time horizon  $[0, t_f]$ , where  $t_f$  denotes the time point at which the bounds are to be evaluated, into  $n_T$  intervals.
- In line with Dowdy and Barton's original work [22], we employ exponential test functions of the form  $g(t) = e^{\rho(t_T - t)}$ . As argued in Example 3.1, any set of test functions of this form satisfies the hypotheses of Propositions 3.2 and 3.3. Throughout, we choose  $t_T$  to coincide with the end of the time horizon on which the bounds are to be evaluated. As  $t_T$  merely controls the scale of the generalized moments generated by  $g$ , this choice is somewhat arbitrary but contributes in our



experience to improved numerical conditioning of problem (SDP). For the choice of the parameters  $\rho$ , we draw motivation from linear time-invariant systems theory and choose  $\rho$  based on the singular values of the coefficient matrix  $A$  of the moment dynamics (mCME). Overall, we choose the test function set  $F = \{e^{-\sigma_i(t_T-t)}\}_{i=1}^{n_F}$  assembled from the smallest  $n_F$  unique singular values  $\sigma_1, \dots, \sigma_{n_F}$  of  $A$ .

- Motivated by the scaling of the size of the bounding SDP (see Table 1), we use the following greedy procedure to ultimately choose  $m, n_I, n_T$  and  $n_F$ 
  1. Fix  $m = 2, n_I = 2, n_F = 1$  and successively increase  $n_T$  until no significant bound tightening effect is observed.
  2. Increase  $n_F$  successively until no significant bound tightening effect is observed.
  3. Increase  $m$  until bounds are sufficiently tight or the computational cost exceeds a tolerable amount.

Note that the above procedure fixes the hierarchy level  $n_I$  at 2. While increasing  $n_I$  generally also has a bound tightening effect, in our experience it promotes numerical ill-conditioning and is rarely significantly more efficient than the other bound tightening mechanisms.

We study the bound tightening effect of the different hierarchy parameters in greater detail with an example in Sect. 5.2. In all other case studies, we employ the above heuristics to choose the hierarchy parameters. As the time point and test function sets are systematically generated once the parameters  $n_T$  and  $n_F$  are chosen, we instead report these parameters in place of T and F throughout.

### 5.1.4 Numerical Considerations

Sum-of-squares and moment problems are notorious for poor numerical conditioning, and the Problems (SDP) and (pwpOCP) are no exception to this issue. And while promising general remedies, such as the use of better suited polynomial bases like Chebyshev polynomials [34], exist in theory, it remains a largely open research problem how to deploy them effectively in practice. To circumvent this deficiency, we instead employ a simple scaling strategy for the decision variables in the bounding problems. This strategy is applicable whenever bounds are computed at multiple time points  $t_1 < t_2 < \dots < t_T$  along a trajectory and can be summarized as follows: we solve the SDPs in chronological order and scale the decision variables in the bounding problem corresponding to the time point  $t_k$  by the values attained at the solution of the bounding problem associated with the previous time point  $t_{k-1}$ . For the initial problem, we scale the system state such that the moments of the initial distribution lie between 0 and 1. This strategy defers the orders of magnitude of the moments to the coefficients in the constraints of Problem (SDP) and as such exposes them directly to the SDP solver which is greatly beneficial in our experience.

While an appropriate scaling of the decision variables is crucial to avoid numerical issues, it is not always sufficient to achieve convergence of the solver to a desired degree of accuracy with respect to optimality. We hedge against potentially inaccurate, suboptimal solutions and ensure validity of the computed bounds by verifying that the solver converged to a dual feasible point and report the associated dual objective value.

### 5.1.5 Implementation

All semidefinite programs solved for the case studies presented in this section were assembled using JuMP [23] and solved with MOSEK v9.0.97 [6]. Our implementation is publicly available at <https://github.com/FHoltorf/StochMP>.

## 5.2 Bound Tightening Mechanisms

In this section, we assess the effect of the different bound tightening mechanisms provided by the proposed bounding hierarchy. We conduct this empirical analysis on the basis of the nonlinear birth-death process



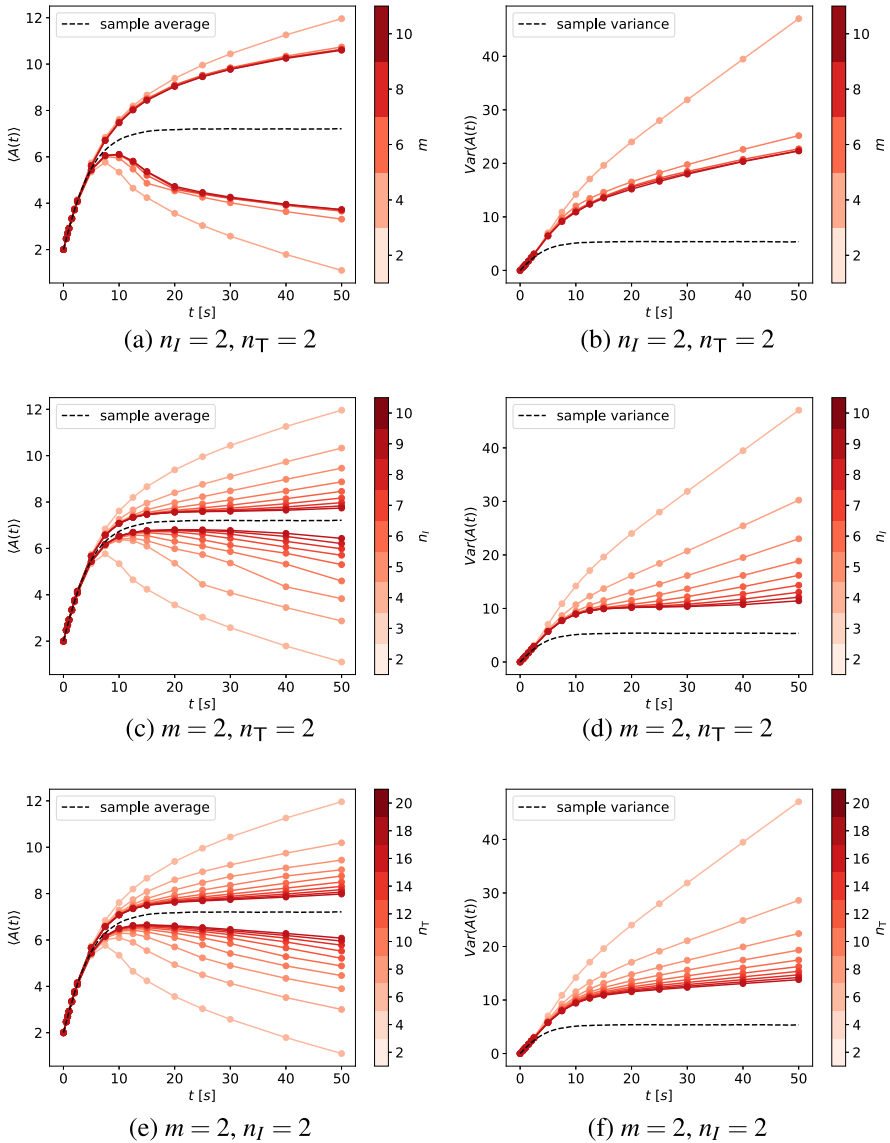
Note that this reaction system has an unbounded state space. Such systems are of particular interest for the application of moment bounding schemes as their moments can rarely be found analytically and the corresponding CME gives rise to a countably infinite system of coupled ODEs precluding a direct numerical integration.

For the sake of simplicity, we restrict our considerations here to studying the bound tightening effect of increasing the truncation order  $m$ , the hierarchy level  $n_I$  and the number of time points  $n_T$  used to discretize the horizon; throughout, we only use the constant test function ( $n_F = 1$ ).

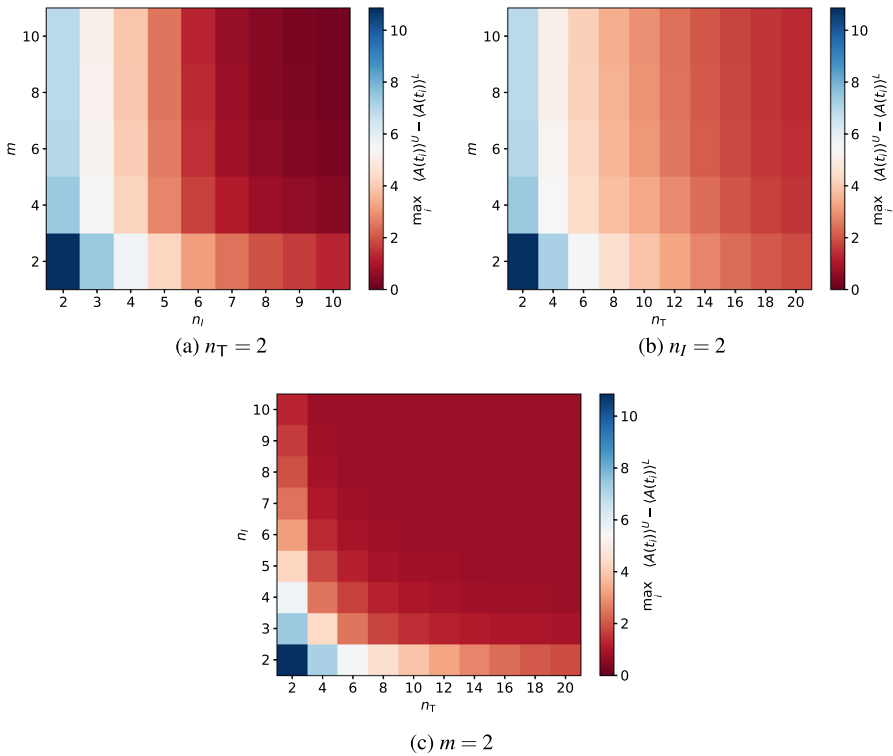
Figure 1 shows the effect of isolated changes in the different hierarchy parameters on the bounds obtained for the mean molecular count of  $A$  and its variance. The results indicate that all bound tightening mechanisms, when used in isolation, appear to suffer from diminishing returns, eventually causing the bounds to stall. Moreover, solely increasing the truncation order  $m$  appears insufficient to provide informative bounds over a long time horizon in this example; increasing either the number of time points or the hierarchy level  $n_I$  is significantly more effective in comparison.

Figure 2 shows the effect of joint changes in the considered hierarchy parameters on the tightness of bounds on the mean molecular count. The figure indicates that jointly changing the hierarchy parameters effectively mitigates stalling of the bounds in this example such that significantly tighter bounds are obtained overall. While the general trends illustrated in Figs. 1 and 2 align well with our experiences for a range of other examples, we wish to emphasize that it is in general hard to predict which combination of hierarchy parameters provides the best trade-off between computational cost and bound quality; when the choice of test functions is added to the equation, the situation becomes even more complicated. Moreover, as Fig. 3 illustrates, the feasible region of the bounding SDPs shrinks anisotropically and, more importantly, with different intensity along different directions for different bound tightening mechanisms. This indicates that the optimal choice of the hierarchy parameters is in general not only dependent on the system under investigation but also on the statistical quantity to be bounded.

In summary, the results presented in this section underline the value of the additional bound tightening mechanisms offered by the proposed hierarchy; however, they also



**Fig. 1** Bounds on the trajectories of the mean molecular count and variance of the birth-death process (14) for increasing  $m$  (a, b),  $n_I$  (c, d), and  $n_T$  (e, f) compared against the empiric sample mean and variance generated with Gillespie’s Stochastic Simulation Algorithm. In each figure only one parameter is varied, while the others are held constant at the level indicated in the subcaptions



**Fig. 2** Maximum gap between upper and lower bounds on mean molecular count of the birth-death process (14) among the time points probed along the time horizon for joint changes in  $m$  and  $n_I$  (a),  $m$  and  $n_T$  (b), and  $n_I$  and  $n_T$  (c)

emphasize the need for better guidelines to enable an effective use of the tightening mechanisms in practice.

### 5.3 Generic Examples

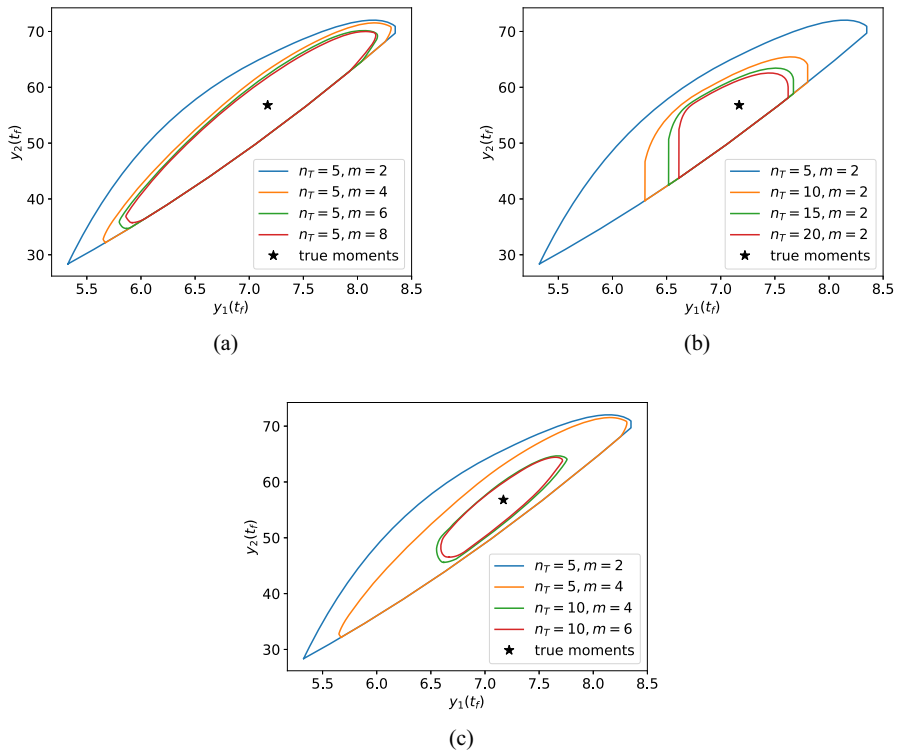
To contrast the performance of the proposed methodology with its predecessor, we first consider two generic reaction networks that were studied by Dowdy and Barton [22].

#### 5.3.1 Simple Reaction Network

First, we study the bound quality for means and variances of the molecular counts of the species A and C following the simple reaction network



Figure 4 shows a comparison between the bounds obtained by both methods. For reference, also the trajectories obtained with Gillespie’s Stochastic Simulation Algorithm



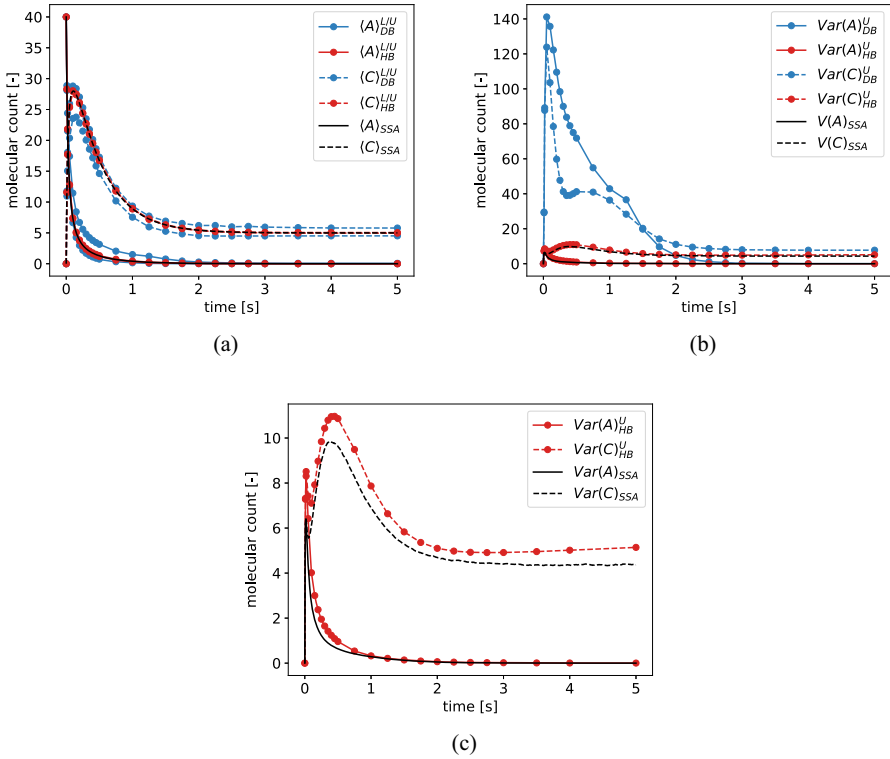
**Fig. 3** Projection of the feasible set of (SDP) corresponding to the birth-death process (14) for (a) increasing truncation orders  $m$ , (b) increasing number of time points  $n_T$ , and (c) jointly increasing truncation orders and number of time points. All projections are obtained for  $n_I = 2$  and  $t_f = 20$

are provided. The results showcase that the presented necessary moment conditions have the potential to tighten the obtained bounds significantly. In particular bounds on the variance of both species are dramatically improved at the relatively low truncation order of  $m = 4$ .

### 5.3.2 Large Reaction Network

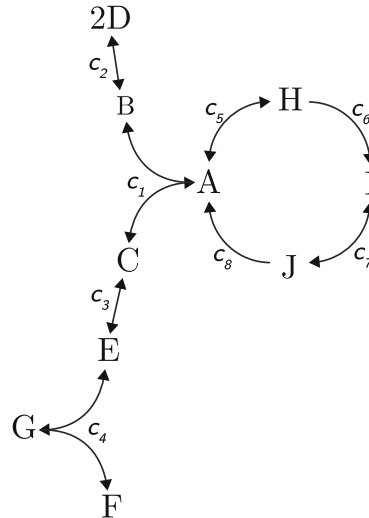
Departing from examples of toyish size, we now study the reaction network shown in Fig. 5. In contrast to the previous reaction systems, this large reaction network poses a major challenge for sampling-based analysis techniques. The underlying reason for that is twofold. On the one hand, the network is characterized by a large, 7-dimensional state space containing hundreds of millions of reachable states [22]. On the other hand, the system is extremely stiff. These properties frustrate sampling-based techniques as they exacerbate the need for large sample sizes and render each sample path evaluation expensive.

Figure 6 shows bounds on the mean molecular counts of species A and H. In line with the results of the previous sections, the bounds obtained by the proposed method are again considerably tighter. In this example, however, this result carries more weight

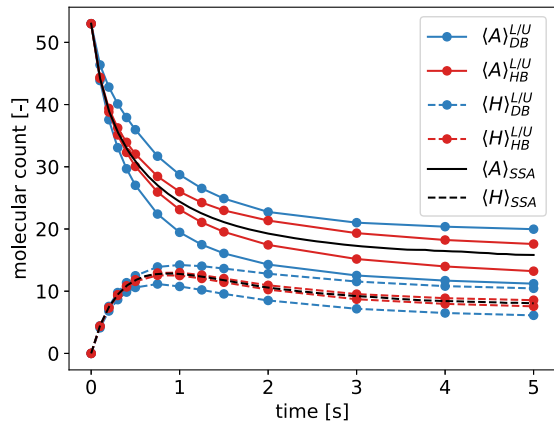


**Fig. 4** Bounds on (a) means and (b, c) variances of molecular counts of species A and C in reaction network (15); initial state:  $x_{A,0} = 40, x_{B,0} = 41,$  and  $x_{C,0} = x_{D,0} = 0$ ; kinetic parameters:  $c = (1, 2.1, 0.3) s^{-1}$ ; hierarchy parameters:  $m = 4, n_F = 3, n_T = 10$

**Fig. 5** Large reaction network from Dowdy and Barton [22]



**Fig. 6** Bounds on the mean molecular counts of species A and H in the large reaction network shown in Fig. 5; initial state:  $x_{A,0} = x_{F,0} = 53$ ,  $x_{B,0} = x_{C,0} = x_{D,0} = x_{E,0} = x_{G,0} = x_{H,0} = x_{I,0} = x_{J,0} = 0$ ; kinetic parameters:  $c = (1, 1, 1, 10^4, 1, 1, 10^5, 1) \text{ s}^{-1}$ ; hierarchy parameters:  $m = 2$ ,  $n_F = 5, n_T = 5$



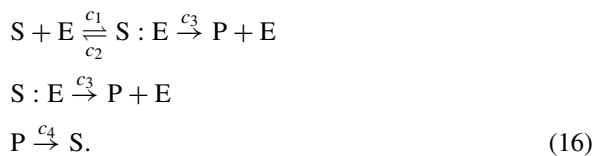
as increasing the truncation order leads to a prohibitive increase in problem size for the method of Dowdy and Barton [22]. Accordingly, the proposed bounding scheme offers bounds at a quality that was previously not attainable for problems of such complexity.

### 5.4 Biochemical Reaction Networks

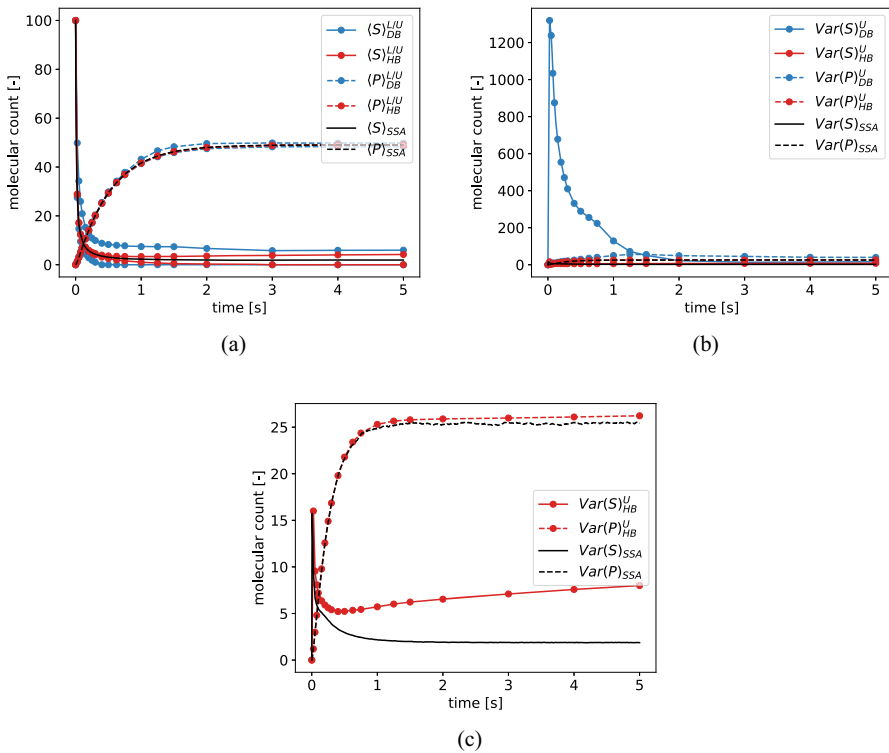
To finally demonstrate that the proposed methodology may in fact be useful in practice, we examine two reaction networks drawn from biochemical applications. In these applications, molecular counts are often in the order of 10s to 100s necessitating the consideration of stochasticity.

#### 5.4.1 Michaelis–Menten Kinetics

Michaelis–Menten kinetics underpin a vast range of metabolic processes. Understanding the behavior and noise present in the associated reaction networks is of particular value for the investigation of the metabolic degradation of trace substances in biological organisms. We examine the basic Michaelis–Menten reaction network:



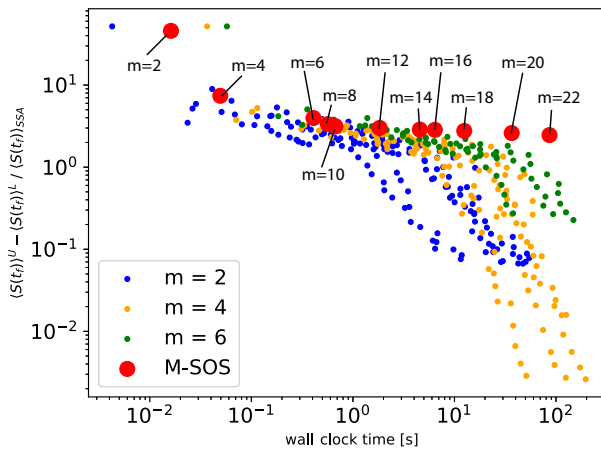
Due to reaction invariants, the reaction network is characterized entirely by a two-dimensional state space. Accordingly, we bound the means and variances of the molecular counts of only the product P and substrate S. The results are illustrated in Fig. 7. The proposed method produces high-quality bounds for the means while providing informatively tight bounds for the variances. Further, it again outperforms its predecessor, especially with respect to the bounds on the variance.



**Fig. 7** Bounds on means (a) and variances (b, c) of the molecular counts of species S and P in the metabolic reaction network (16); initial state:  $x_{S,0} = x_{E,0} = 100$ ,  $x_{P,0} = x_{S;E,0} = 0$ ; kinetic parameters:  $\mathbf{c} = (1, 1, 1, 1) \text{ s}^{-1}$ ; hierarchy parameters:  $m = 4$ ,  $n_F = 3$ ,  $n_T = 20$

To showcase further that the proposed hierarchy can provide notable computational savings in practice, we also investigate the trade-off between computational cost and bound quality for a wide range of hierarchy parameters. As the standard moment-sum-of-squares hierarchy, i.e., the use of monomial test functions, provides an established alternative for bounding moment trajectories of stochastic chemical systems [61], we compare against this baseline. Figure 8 shows this comparison for bounds on the mean molecular count of species S at the end of the time horizon  $t_f = 5 \text{ s}$ . The proposed bounding problems allow for the computation of overall tighter bounds, often at a dramatic cost reduction. This advantage persists for almost all choices of the hierarchy parameters from the tested range, albeit some choices are significantly better than others. The results further indicate that the new bound tightening mechanisms of refining the time domain partition and increasing the hierarchy level  $n_I$  can be more effective than increasing the truncation order  $m$ , emphasizing the practical improvement upon Dowdy and Barton’s method.





**Fig. 8** Trade-off between bound quality and computational cost for the proposed hierarchy and the standard moment-sum-of-squares hierarchy when applied to the metabolic reaction network (16). Small markers refer to results obtained with the proposed bounding problems. For the proposed bounding problems, instances were constructed using hierarchy parameters in the range  $m \in \{2, 4, 6\}$ ,  $n_T \in \{1, \dots, 100\}$ ,  $n_I \in \{1, \dots, 7\}$  and  $n_R \in \{1, 2, 3\}$  so that the computation time does not exceed 200s. For the moment sum-of-squares hierarchy (M-SOS), the total degree  $m$  of monomial test functions was varied

### 5.4.2 Negative Feedback Biocircuit

Many efforts of modern synthetic biology culminate in the design of biocircuits subject to stringent constraints on robustness and performance. Upon successful design, the implications of such tailored biocircuits are often far reaching, even addressing global challenges such as water pollution [66] and energy [55]. Accordingly, in recent years the use of systems theoretic techniques to conceptualize and speed up the design process of biocircuits has received considerable attention [19]. In this context, Sakurai and Hori [60] demonstrated the utility of bounds on stationary moments for the design of biocircuits subject to robustness constraints.

We show here that the proposed bounding method could enable an extension of Sakurai and Hori’s analysis to the dynamic case. To that end, we consider the same negative feedback biocircuit as studied in [60]. The reaction network is illustrated in Fig. 9 and formally given by the following set of reactions

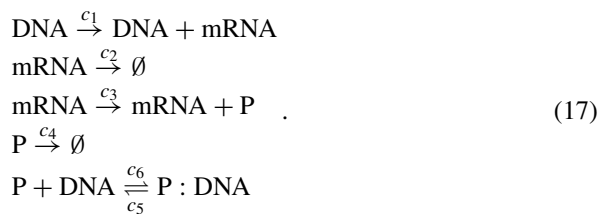
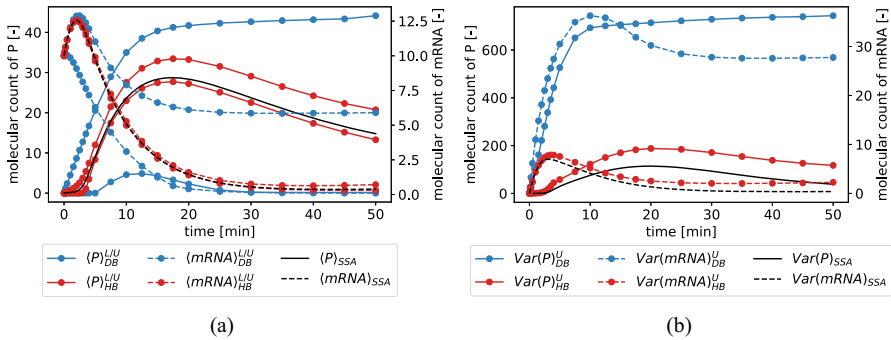
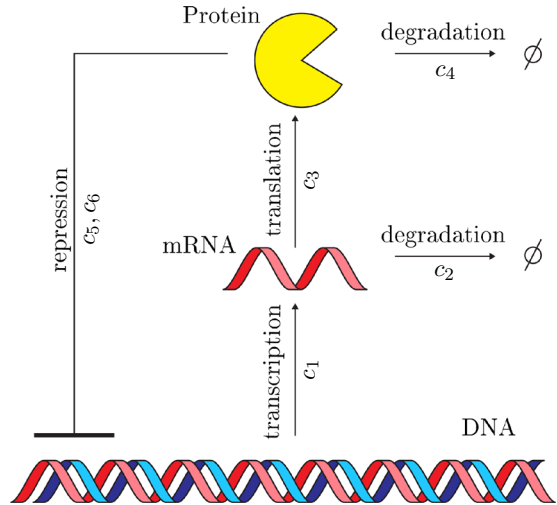


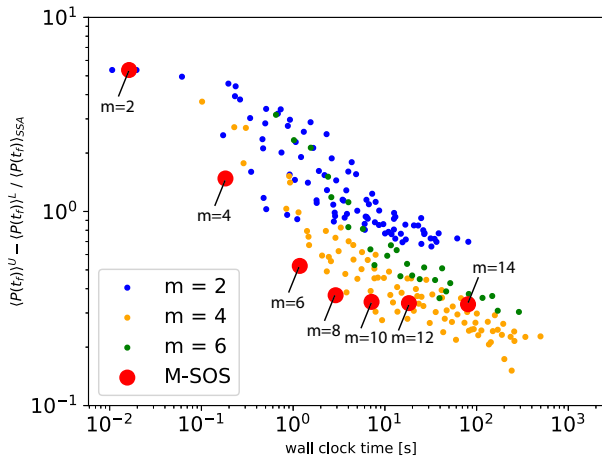
Figure 10 illustrates the obtained bounds on means and variances of the molecular counts of mRNA and protein (P). The bounds are of high quality and may provide

**Fig. 9** Negative feedback biocircuit. Reproduced from Sakurai and Hori [60]



**Fig. 10** Bounds on means (a) and variances (b) of the molecular counts of mRNA and P in the negative feedback biocircuit illustrated in Fig. 9; initial state:  $x_{\text{mRNA},0} = 10, x_P = x_{P:\text{DNA},0} = 0, x_{\text{DNA},0} = 20$ ; kinetic parameters:  $c = (0.2, \ln(2)/5, 0.5, \ln(2)/20, 5, 1) \text{ min}^{-1}$ ; hierarchy parameters:  $m = 4, n_F = 3, n_T = 20$

useful information for robustness analysis as the noise level measured by the variance changes significantly over the time horizon until a stationary value is reached. Figure 10 also underlines that the proposed bounding scheme is capable of furnishing considerably tighter bounds at the same truncation order as its predecessor. Figure 11 further provides a comparison of the bound quality vs. cost trade-off against the baseline of the standard moment-sum-of-squares hierarchy. While the proposed bounding scheme does not outperform the standard moment-sum-of-squares hierarchy at low precision for this example, it enables the computation of tighter bounds at similar cost overall. Most critically, it needs to be emphasized that the bounding problems of the moment-sum-of-squares hierarchy could not be solved beyond the truncation order  $m = 14$  due to numerical issues. This highlights the practical value of the proposed tightening mechanisms that avoid increasing the truncation order.



**Fig. 11** Trade-off between bound quality and computational cost for the proposed hierarchy and the standard moment-sum-of-squares hierarchy when applied to the biocircuit (17). Small markers refer to results obtained with the proposed bounding problems. For the proposed bounding problems, instances were constructed using hierarchy parameters in the range  $m \in \{2, 4, 6\}$ ,  $n_T \in \{1, \dots, 100\}$ ,  $n_I \in \{1, \dots, 7\}$  and  $n_R \in \{1, 2, 3\}$  so that computation time does not exceed 1000s. For the moment sum-of-squares hierarchy (M-SOS), the total degree  $m$  of monomial test functions was varied

## 6 Conclusion

### 6.1 Summary

We have extended the results of Dowdy and Barton [22] by constructing a new hierarchy of convex necessary moment conditions for the moment trajectories of stochastic chemical systems described by the CME. Building on a discretization of the time domain of the problem, the conditions reflect temporal causality and regularity properties of the true moment trajectories. It is proved that the conditions give rise to a hierarchy of SDPs whose optimal values form a sequence of monotonically improving bounds on the true moment trajectories. Furthermore, the conditions provide new mechanisms to tighten the obtained bounds when compared to the original conditions proposed by Dowdy and Barton [22]. These tightening mechanisms are often a more scalable and practical alternative to the primary tightening mechanism of increasing the moment truncation order in Dowdy and Barton’s moment bounding scheme [22]; most notably, refining the time discretization results in linearly increasing problem sizes, independent of the state space dimension of the system. As an additional advantage, this bound tightening mechanism provides a way to sidestep the poor numerical conditioning of moment-based SDPs featuring high-order moments. Finally, it is demonstrated with several examples that the proposed hierarchy provides bounds that may indeed be useful in practice.

## 6.2 Open Questions

We close by stating some open questions motivated by our results.

1. In the presented case studies, we naively chose the time points at which the proposed necessary moment conditions were imposed as equidistant. Several results from numerical integration, perhaps most notably Gauß quadrature, suggest that this is likely a suboptimal choice. It would be interesting to examine if and how results from numerical integration can inform improvements of this choice.
2. The choice of the hierarchy parameters in the proposed bounding scheme is crucial to achieve a good trade-off between bound quality and computational cost. As indicated by the discussion in Sect. 5.2, however, the interplay between the bound tightening mechanisms associated with the different hierarchy parameters and their effect on the bound quality remains poorly understood. Accordingly, we believe that assessing the trade-offs offered by the different bound tightening mechanisms in greater detail and developing more rigorous guidelines on how to utilize them effectively constitutes an important step toward improving the practicality of the proposed method.
3. The ideas discussed in Sect. 4 constitute additional promising research avenues toward improving practicality of the proposed method. Specifically, there are several open questions pertaining to the concrete implementation of Algorithm 1 and the way Problem (pwpOCP) can be used to inform an effective use of the different bound tightening mechanisms. Furthermore, the decomposable, weakly coupled structure of the bounding SDPs motivates other forms of exploitation than Algorithm 1; in particular the use of distributed optimization techniques such as ADMM [14] or overlapping Schwarz decomposition [51, 65] appears promising.

**Funding** Open Access funding provided by the MIT Libraries.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Proof of Proposition 3.1

**Proof** We will first introduce a convenient formalism to describe the conditions in the form of Eq. (7) in terms of the coefficient matrix  $\mathbf{A}$  of the moment dynamics

$$\mathbf{K} \frac{d\mathbf{y}}{dt}(t) = \mathbf{A}\mathbf{y}(t).$$

To that end, we assume that the moments are ordered according to their corresponding multi-indices, i.e., the first entry corresponds to  $\mathbf{y}_0$ , the following  $n$  entries correspond

to the moments  $y_j$  with  $|j| \leq 1$  and so on. Then, we define the matrices  $\mathbf{K}_l = [\mathbf{I}_{n_L(l) \times n_L(l)} \ \mathbf{0}_{n_L(l) \times n_H(l)}]$  with  $n_L(l) = \binom{m-lq+n}{n}$  and  $n_H(l) = \binom{m+q+n}{n} - n_L(l)$ . In words, applying  $\mathbf{K}_l$  to  $\mathbf{y}$  extracts all moments up to order  $m-lq$ . Now we can express the higher derivatives  $\frac{d^k}{dt^k}(\mathbf{K}_l \mathbf{y})$  for any  $k \leq l+1$  in the following way: let  $\mathbf{A}_i$  be the coefficient matrix such that  $\frac{d}{dt}(\mathbf{K}_{i+1} \mathbf{y}) = \mathbf{A}_i \mathbf{K}_i \mathbf{y}$ . By induction, we then arrive at the convenient expression

$$\frac{d^k}{dt^k}(\mathbf{K}_l \mathbf{y}) = \prod_{i=1}^k \mathbf{A}_{l-i} \mathbf{K}_{l-k} \mathbf{y}.$$

Recalling the definition of Eq. (7) in terms of the Taylor series expansion of the true moment trajectories, it follows that the equation system

$$\mathbf{K}_l \mathbf{y}_{t_2} = \sum_{k=0}^l \frac{(t_2 - t_1)^k}{k!} \prod_{i=1}^k \mathbf{A}_{l-i} \mathbf{K}_{l-k} \mathbf{y}_{t_1} + \frac{1}{l!} \prod_{i=1}^{l+1} \mathbf{A}_{l-i} \mathbf{z}_{g_l, t_2} \tag{16_l}$$

corresponds to all equations of the form of Eq. (7) with  $|j| \leq m-lq$  for fixed  $l \in \{0, \dots, n_I\}$ . Note that we already used the fact that  $\mathbf{K}_{-1}$  reduces to the identity in the above relation.

Now notice that Eq. (6) holds for all  $l = 0, \dots, n_I$  if and only if

$$\mathbf{K}(\mathbf{y}_{t_2} - \mathbf{y}_{t_1}) = \mathbf{A} \mathbf{z}_{g_0, t_2} \tag{18}$$

holds alongside

$$\frac{1}{(l-1)!} \prod_{i=1}^l \mathbf{A}_{l-i} \mathbf{K}_0 \mathbf{z}_{g_{l-1}, t_2} = \frac{(t_2 - t_1)^l}{l!} \prod_{i=1}^l \mathbf{A}_{l-i} \mathbf{K}_0 \mathbf{y}_{t_1} + \frac{1}{l!} \prod_{i=1}^{l+1} \mathbf{A}_{l-i} \mathbf{z}_{g_l, t_2} \tag{19}$$

for  $l = 1, \dots, n_I$  since Eq. (18) is equivalent to Eqs. (16<sub>0</sub>) and Eq. (19) is obtained simply by matching coefficients of Eqs. (16<sub>l</sub>) and (16<sub>l-1</sub>) for each  $l \in \{1, \dots, n_I\}$ .

Factoring out  $\prod_{i=1}^l \mathbf{A}_{l-i}$  in Eq. (19) and noting that  $\mathbf{A}_{-1} = \mathbf{A}$  as well as  $\mathbf{K}_0 = \mathbf{K}$  shows that Eq. (19) is implied by

$$l \mathbf{K} \mathbf{z}_{g_{l-1}, t_2} = (t_2 - t_1)^l \mathbf{K} \mathbf{y}_{t_1} + \mathbf{A} \mathbf{z}_{g_l, t_2}. \tag{20}$$

Thus, by using that  $g_l(t_2) = 0$  for  $l = 1, \dots, n_I$ , it follows that Eq. (20) is equivalent to Eq. (6) for  $l = 1, \dots, n_I$ . We finally conclude the proof by noting that Eq. (18) is equivalent to Eq. (6) for  $l = 0$ . □

### Appendix B: Commutative Property of $I_L$ and $I_R$

**Lemma B.1** *The left and right integral operators  $I_L, I_R : C(\mathbb{R}^2) \rightarrow C(\mathbb{R}^2)$  defined by*

$$(I_L f)(t_1, t_2) = \int_{t_1}^{t_2} f(t_1, t) dt$$

$$(I_R f)(t_1, t_2) = \int_{t_1}^{t_2} f(t, t_2) dt.$$

*commute.*

**Proof** Using the definition of  $I_L$  and  $I_R$  we obtain

$$(I_L I_R f)(t_1, t_2) = \int_{t_1}^{t_2} (I_R f)(t_1, t) dt = \int_{t_1}^{t_2} \int_{t_1}^t f(s, t) ds dt.$$

By defining  $\Omega_{LR} = \{(s, t) \in \mathbb{R}^2 : t_1 \leq s \leq t, t_1 \leq t \leq t_2\}$  we can write more concisely

$$(I_L I_R f)(t_1, t_2) = \int_{\Omega_{LR}} f(s, t) d(s, t).$$

Likewise, we have that

$$(I_R I_L f)(t_1, t_2) = \int_{t_1}^{t_2} (I_L f)(s, t_2) ds = \int_{t_1}^{t_2} \int_s^{t_2} f(s, t) dt ds = \int_{\Omega_{RL}} f(s, t) d(s, t),$$

where  $\Omega_{RL} = \{(s, t) \in \mathbb{R}^2 : t_1 \leq s \leq t_2, s \leq t \leq t_2\}$ . Observing that  $\Omega_{RL} = \Omega_{LR}$  concludes the proof. □

### Appendix C: Proof of Corollary 3.1

**Proof** Let  $l \in \{0, \dots, n_I\}$  be fixed. We show that

$$z(h_l g; t_2) = l! \left( z^{l+1}(g; t_2) - \sum_{k=0}^l \frac{(t_2 - t_1)^{l-k}}{(l-k)!} z^{k+1}(g; t_1) \right) \tag{21}$$

satisfies the desired conditions. First note that Eq. (21) is equivalent to

$$z(h_l g; t_2) = l! \Omega_{l+1,0} \left( \left\{ z^i(g; t_1) \right\}_{i=1}^{l+1}, \left\{ z^i(g; t_2) \right\}_{i=1}^{l+1}, t_1, t_2 \right).$$

Since by assumption

$$\Omega_{l+1,0} \left( \left\{ z^i(g; t_1) \right\}_{i=1}^{l+1}, \left\{ z^i(g; t_2) \right\}_{i=1}^{l+1}, t_1, t_2 \right) \in C(X),$$

it follows that Condition (10) holds for the definition in Eq. (21). Next, note that Condition (10) can be equivalently written as

$$Az(h_l g; t_2) = K \left( lz(h_{l-1} g; t_2) - z(h_l g'; t_2) - (t_2 - t_1)^l g(t_1) y(t_1) \right)$$

using the definition of  $h_l$  and that by assumption  $z((h_l g)'; t_2) = z(h_l' g; t_2) + z(h_l g'; t_2)$ . Thus, we need to show that the following identity holds:

$$\begin{aligned} & A \left( z^{l+1}(g; t_2) - \sum_{k=0}^l \frac{(t_2 - t_1)^{l-k}}{(l-k)!} z^{k+1}(g; t_1) \right) \\ &= K \left( z^l(g; t_2) - \sum_{k=0}^{l-1} \frac{(t_2 - t_1)^{l-1-k}}{(l-1-k)!} z^{k+1}(g; t_1) \right) \\ &\quad - K \left( z^{l+1}(g'; t_2) - \sum_{k=0}^l \frac{(t_2 - t_1)^{l-k}}{(l-k)!} z^{k+1}(g'; t_1) \right) - K \frac{(t_2 - t_1)^l}{l!} g(t_1) y(t_1). \end{aligned}$$

By rearranging this equation as

$$\begin{aligned} & \sum_{k=0}^l \frac{(t_2 - t_1)^{l-k}}{(l-k)!} \left( A \left[ z^{k+1}(g; t_2) - z^{k+1}(g; t_1) \right] + K \left[ z^{k+1}(g'; t_2) - z^{k+1}(g'; t_1) \right] \right) \\ & \quad - \sum_{k=1}^l \frac{(t_2 - t_1)^{l-k}}{(l-k)!} K \left( z^k(g; t_2) - z^k(g; t_1) \right) \\ &= \sum_{k=0}^{l-1} \frac{(t_2 - t_1)^{l-k}}{(l-k)!} \left( Az^{k+1}(g; t_2) + Kz^{k+1}(g'; t_2) \right) - \sum_{k=1}^{l-1} \frac{(t_2 - t_1)^{l-k}}{(l-k)!} Kz^k(g; t_2) \\ & \quad - \frac{(t_2 - t_1)^l}{l!} K g(t_1) y(t_1) \end{aligned}$$

and using that

$$Az^{k+1}(g; t_i) + Kz^{k+1}(g'; t_i) = K \left( z^k(g; t_i) - \frac{t_i^k}{k!} g(0) y(0) \right)$$

holds for all  $k \in \{0, \dots, n_l\}$  and  $i \in \{1, 2\}$  with  $z^0(g; t_i) = g(t_i) y(t_i)$ , we obtain after simplifying and collecting terms that

$$\begin{aligned} & \sum_{k=1}^l \frac{(t_2 - t_1)^{l-k} (t_1^k - t_2^k)}{(l-k)! k!} K g(0) y(0) + \frac{(t_2 - t_1)^l}{l!} K (g(t_2) y(t_2) - g(t_1) y(t_1)) \\ &= - \sum_{k=0}^{l-1} \frac{(t_2 - t_1)^{l-k} t_2^k}{(l-k)! k!} K g(0) y(0) + \frac{(t_2 - t_1)^l}{l!} K (g(t_2) y(t_2) - g(t_1) y(t_1)). \end{aligned}$$

Applying the binomial theorem finally verifies that this equation is true, so Condition (10) is indeed satisfied for  $z(h_l g; t_2)$  as defined in Eq. (21). □

### Appendix D: Proof of Corollary 3.2

Our proof of Corollary 3.2 relies in large part on explicit computations. To that end, it is first essential to provide an explicit algebraic relation for  $\Omega_{l,k} \left( \{z^s(g; t_1)\}_{s=1}^l, \{z^s(g; t_2)\}_{s=1}^l, t_1, t_2 \right)$ . Recall that  $\Omega_{k,l}$  was introduced to compactly denote expressions of the form  $(I_L^n I_R^m f)(t_1, t_2)$  with  $f(x, y) = z^1(g; y) - z^1(g; x)$ . To avoid unnecessary clutter of notation, we will give an explicit algebraic expression for the latter; adjusting the indices to obtain an expression for  $\Omega_{l,k}$  is straightforward and can be found in the proof of Corollary 3.2. It is easily verified by induction that for any  $m, n \in \mathbb{Z}_+$ , the following identity holds:

$$\begin{aligned} (I_L^n I_R^m f)(t_1, t_2) &= \sum_{k=0}^m (-1)^k \binom{n+k}{n} \frac{(t_2 - t_1)^{m-k}}{(m-k)!} z^{n+1+k}(g; t_2) \\ &\quad + (-1)^{m+1} \sum_{k=0}^n \binom{m+k}{m} \frac{(t_2 - t_1)^{n-k}}{(n-k)!} z^{m+1+k}(g; t_1). \end{aligned} \tag{22}$$

With this in hand, we will first proof the following Lemma which answers the essential question behind Corollary 3.2.

**Lemma D.1** *Let  $z^l$  be a sequence of functions that satisfies  $z^l(t) = \int_0^t z^{l-1}(\tau) d\tau$  for  $l \geq 2$ . Further, define  $f(x, y) = z^1(y) - z^1(x)$  and consider a convex cone  $C$ . Then, for any  $0 \leq t_1 \leq t_2 \leq t_3 < +\infty$  and  $n, m \in \mathbb{Z}_+$ , the following implications hold:*

$$\begin{aligned} (I_L^p I_R^s f)(t_1, t_2), (I_L^p I_R^s f)(t_2, t_3) \in C \text{ for all } 0 \leq p \leq n+1 \text{ and } 0 \leq s \leq m \\ \implies (I_L^{n+1} I_R^m f)(t_1, t_3) \in C \end{aligned} \tag{23}$$

and

$$\begin{aligned} (I_L^p I_R^s f)(t_1, t_2), (I_L^p I_R^s f)(t_2, t_3) \in C \text{ for all } 0 \leq p \leq n \text{ and } 0 \leq s \leq m+1 \\ \implies (I_L^n I_R^{m+1} f)(t_1, t_3) \in C. \end{aligned} \tag{24}$$

**Proof** We focus on proving Implication (23). To that end, note that in order to establish Implication (23), it suffices to show that

$$\int_{t_2}^{t_3} (I_L^n I_R^m f)(t_1, t) - (I_L^n I_R^m f)(t_2, t) dt \in C \tag{25}$$

holds as

$$(I_L^{n+1} I_R^m f)(t_1, t_3) = \int_{t_1}^{t_2} (I_L^n I_R^m f)(t_1, t) dt + \int_{t_2}^{t_3} (I_L^n I_R^m f)(t_1, t) dt.$$

We will thus show that the left-hand side of Eq. (25) can be expressed as a conic combination of  $(I_L^p I_R^s f)(t_1, t_2)$  and  $(I_L^p I_R^s f)(t_2, t_3)$  with  $0 \leq p \leq n+1$  and  $0 \leq s \leq$



$m$ . From the definition of  $I_L$  and  $I_R$ , we obtain

$$\begin{aligned}
 & (I_L^n I_R^m f)(t_1, t) - (I_L^n I_R^m f)(t_2, t) \\
 &= \sum_{k=0}^m (-1)^{m-k} \binom{n+m-k}{n} \frac{(t-t_1)^k}{k!} z^{n+m+1-k}(t) \\
 &\quad - \sum_{k=0}^m (-1)^{m-k} \binom{n+m-k}{n} \frac{(t-t_2)^k}{k!} z^{n+m+1-k}(t) \\
 &\quad + (-1)^{m+1} \sum_{k=0}^n \binom{m+n-k}{m} \frac{(t-t_1)^k}{k!} z^{n+m+1-k}(t_1) \\
 &\quad - (-1)^{m+1} \sum_{k=0}^n \binom{m+n-k}{m} \frac{(t-t_2)^k}{k!} z^{n+m+1-k}(t_2). \tag{26}
 \end{aligned}$$

Using the binomial formula, it further follows that

$$\begin{aligned}
 & \sum_{k=0}^m (-1)^{m-k} \binom{n+m-k}{n} \frac{(t-t_1)^k}{k!} z^{n+m+1-k}(t) \\
 &= \sum_{k=0}^m (-1)^{m-k} \binom{n+m-k}{n} z^{n+m+1-k}(t) \sum_{l=0}^k \frac{(t-t_2)^{k-l}}{(k-l)!} \frac{(t-t_1)^l}{l!} \\
 &= \sum_{l=0}^m \frac{(t_2-t_1)^l}{l!} \sum_{k=0}^{m-l} (-1)^{m-k-l} \binom{n+m-k-l}{n} \frac{(t-t_2)^k}{k!} z^{n+m+1-k-l}(t)
 \end{aligned}$$

and likewise

$$\begin{aligned}
 & (-1)^{m+1} \sum_{k=0}^n \binom{m+n-k}{m} \frac{(t-t_1)^k}{k!} z^{n+m+1-k}(t_1) \\
 &= (-1)^{m+1} \sum_{k=0}^n \binom{m+n-k}{m} z^{n+m+1-k}(t_1) \sum_{l=0}^k \frac{(t_2-t_1)^{k-l}}{(k-l)!} \frac{(t-t_2)^l}{l!} \\
 &= (-1)^{m+1} \sum_{l=0}^n \frac{(t-t_2)^l}{l!} \sum_{k=0}^{n-l} \binom{m+n-k-l}{m} \frac{(t_2-t_1)^k}{k!} z^{n+m+1-k-l}(t_1).
 \end{aligned}$$

Combining the above identities with Eq. (26) yields

$$\begin{aligned}
 & (I_L^n I_R^m f)(t_1, t) - (I_L^n I_R^m f)(t_2, t) \\
 &= \sum_{l=1}^m \frac{(t_2-t_1)^l}{l!} \sum_{k=0}^{m-l} (-1)^{m-k-l} \binom{n+m-l-k}{n} \frac{(t-t_2)^k}{k!} z^{n+m+1-l-k}(t)
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{l=0}^n \frac{(t-t_2)^l}{l!} \left\{ (-1)^{m+1} \sum_{k=0}^{n-l} \binom{m+n-k-l}{m} \frac{(t_2-t_1)^k}{k!} z^{n+m+1-k-l}(t_1) \right. \\
 & \left. - (-1)^{m+1} \binom{m+n-l}{m} z^{n+m+1-l}(t_2) \right\}.
 \end{aligned}$$

By adding and subtracting

$$\begin{aligned}
 & \sum_{l=1}^m (-1)^{m-l+1} \frac{(t_2-t_1)^l}{l!} \sum_{k=0}^n \binom{n+m-l-k}{m-l} \frac{(t-t_2)^k}{k!} z^{n+m+1-l-k}(t_2) \\
 & = \sum_{l=0}^n \frac{(t-t_2)^l}{l!} \sum_{k=1}^m (-1)^{m-k+1} \binom{n+m-l-k}{n-l} \frac{(t_2-t_1)^k}{k!} z^{n+m+1-l-k}(t_2),
 \end{aligned}$$

we get

$$\begin{aligned}
 & (I_L^n I_R^m f)(t_1, t) - (I_L^n I_R^m f)(t_2, t) \\
 & = \sum_{l=1}^m \frac{(t_2-t_1)^l}{l!} \left\{ (-1)^{m-l+1} \sum_{k=0}^n \binom{n+m-l-k}{m-l} \frac{(t-t_2)^k}{k!} z^{n+m+1-l-k}(t_2) \right. \\
 & \quad \left. + \sum_{k=0}^{m-l} (-1)^{m-k-l} \binom{n+m-l-k}{n} \frac{(t-t_2)^k}{k!} z^{n+m+1-l-k}(t) \right\} \\
 & \quad + (-1)^{m+1} \sum_{l=0}^n \frac{(t-t_2)^l}{l!} \left\{ \sum_{k=0}^{n-l} \binom{m+n-k-l}{m} \frac{(t_2-t_1)^k}{k!} z^{n+m+1-k-l}(t_1) \right. \\
 & \quad \left. - \binom{m+n-l}{m} z^{n+m+1-l}(t_2) \right. \\
 & \quad \left. - \sum_{k=1}^m (-1)^{-k} \binom{n+m-l-k}{m-k} \frac{(t_2-t_1)^k}{k!} z^{n+m+1-l-k}(t_2) \right\}
 \end{aligned}$$

which can finally be simplified to

$$\begin{aligned}
 & \sum_{l=1}^m \frac{(t_2-t_1)^l}{l!} \left\{ (-1)^{m-l+1} \sum_{k=0}^n \binom{n+m-l-k}{m-l} \frac{(t-t_2)^k}{k!} z^{n+m+1-l-k}(t_2) \right. \\
 & \quad \left. + \sum_{k=0}^{m-l} (-1)^{m-k-l} \binom{n+m-l-k}{n} \frac{(t-t_2)^k}{k!} z^{n+m+1-l-k}(t) \right\} \\
 & \quad + \sum_{l=0}^n \frac{(t-t_2)^l}{l!} \left\{ (-1)^{m+1} \sum_{k=0}^{n-l} \binom{m+n-k-l}{m} \frac{(t_2-t_1)^k}{k!} z^{n+m+1-k-l}(t_1) \right. \\
 & \quad \left. + \sum_{k=0}^m (-1)^{m-k} \binom{n+m-l-k}{m-k} \frac{(t_2-t_1)^k}{k!} z^{n+m+1-l-k}(t_2) \right\}.
 \end{aligned}$$

Close inspection reveals that this is precisely what we set out to show: every term in the first sum is equivalent to  $(I_L^n I_R^{m-l} f)(t_2, t)$ , while every term in the second sum is equivalent to  $(I_L^{n-l} I_R^m f)(t_1, t_2)$ . Thus, integrating the above relation over  $[t_2, t_3]$  yields a conic combination of  $(I_L^p I_R^s f)(t_1, t_2)$  and  $(I_L^l I_R^s f)(t_2, t_3)$  with  $0 \leq p \leq n + 1$  and  $0 \leq s \leq m$ .

The proof of implication (24) is analogous. □

**Proof of Corollary 3.2** As argued in the beginning of this section,  $\Omega_{l,k}(\{z^s(t_i)\}_{s=1}^l, \{z^s(t_{i+1})\}_{s=1}^l, t_i, t_{i+1})$  takes the following algebraic form

$$\begin{aligned} &\Omega_{l,k} \left( \{z^s(t_i)\}_{s=1}^l, \{z^s(t_{i+1})\}_{s=1}^l, t_i, t_{i+1} \right) \\ &= \sum_{s=0}^k (-1)^s \binom{l-1-k+s}{l-1-k} \frac{(t_2 - t_1)^s}{s!} z^{l-k+s}(t_{i+1}) \\ &\quad + (-1)^{k+1} \sum_{s=0}^{l-1-k} \binom{k+s}{k} \frac{(t_2 - t_1)^{l-1-k-s}}{(l-1-k-s)!} z^{k+1+s}(t_i). \end{aligned}$$

The result therefore follows immediately from Lemma D.1. □

### Appendix E: Reformulation of Condition (i) in Proposition 3.2

In the definition of  $S(F, T, n_I)$  we reformulated Condition (i) of Proposition 3.2 in order to endow  $S(F, T, n_I)$  with a decomposable structure. Without modification, Condition (i) in Proposition 3.2 gives rise to the following necessary moment conditions

$$\Gamma_g \left( \{z^l(f; t)\}_{f \in F} \right) = K \left( z^{l-1}(g; t) - \frac{t^{l-1}}{(l-1)!} g(0) y_0 \right), \quad \forall t \in T. \tag{27}$$

In order to turn this condition into a form that lends itself to be decomposed as described in Sect. 4.1, we can recognize that the above conditions evaluated for different time points are only linked through the term  $g(0)K y_0$ . To expose the recurrent structure of the constraints, we replace this linkage with one between adjacent time points. To that end, we can rearrange the above equation to obtain

$$g(0)K y_0 = \frac{(l-1)!}{t^{l-1}} \left( K z^{l-1}(g; t) - \Gamma_g \left( \{z^l(f; t)\}_{f \in F} \right) \right), \quad \forall t \in T. \tag{28}$$

We then simply replace the linking term in Eq. (27) by the right-hand side of Eq. (28) evaluated at an appropriately shifted time point. Concretely, for the first time point in  $T$  we retain the unmodified equation

$$\Gamma_g \left( \{z^l(f; t_1)\}_{f \in F} \right) = K \left( z^{l-1}(g; t_1) - \frac{t_1^{l-1}}{(l-1)!} g(0) y_0 \right)$$

and for any other  $t \in T$  we reformulate Eq. (27) by replacing  $g(0)Ky_0$  with the right-hand side of Eq. (28) evaluated at the left adjacent time point  $n(t)$ . This yields the desired relation

$$\begin{aligned} & \Gamma_g \left( \{z^l(f; t)\}_{f \in F} \right) - \left( \frac{t}{n(t)} \right)^{l-1} \Gamma_g \left( \{z^l(f; n(t))\}_{f \in F} \right) \\ &= K \left( z^{l-1}(g; t) - \left( \frac{t}{n(t)} \right)^{l-1} z^{l-1}(g; n(t)) \right), \quad \forall t \in T \setminus \{t_1\}. \end{aligned}$$

## References

- Ahmadi, A.A., Dash, S., Hall, G.: Optimization over structured subsets of positive semidefinite matrices via column generation. *Discret. Optim.* **24**, 129–151 (2017)
- Ahmadi, A.A., El Khadir, B.: Time-varying semidefinite programs. *Math. Oper. Res.* **46**(3), 1054–1080 (2021)
- Ahmadi, A.A., Hall, G.: On the construction of converging hierarchies for polynomial optimization based on certificates of global positivity. *Math. Oper. Res.* **44**(4), 1192–1207 (2019)
- Ahmadi, A.A., Majumdar, A.: DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. *SIAM J. Appl. Algebra Geom.* **3**(2), 193–230 (2019)
- Ale, A., Kirk, P., Stumpf, M.P.H.: A general moment expansion method for stochastic kinetic models. *J. Chem. Phys.* **138**(17), 174101 (2013)
- Andersen, E., Andersen, K.: The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In: Frenk, H., Roos, K., Terlaky, T., Zhang, S. (eds.) *High Performance Optimization*, pp. 197–232. Springer, New York (2000)
- Arkin, A., Ross, J., McAdams, H.H.: Stochastic kinetic analysis of developmental pathway bifurcation in phage  $\lambda$ -infected *Escherichia coli* cells. *Genetics* **149**(4), 1633–1648 (1998)
- Artyomov, M.N., Das, J., Kardar, M., Chakraborty, A.K.: Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. *Proc. Natl. Acad. Sci.* **104**(48), 18958–18963 (2007)
- Ascher, U.M., Petzold, L.R.: *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*, vol. 61. SIAM, Philadelphia (1998)
- Backenköhler, M., Bortolussi, L., Wolf, V.: Bounding first passage times in chemical reaction networks. In: Bertino, E., Gao, W., Steffen, B., Woeginger, G., Yung, M. (eds.) *International Conference on Computational Methods in Systems Biology*, pp. 379–382. Springer, New York (2019)
- Bertsimas, D., Caramanis, C.: Bounds on linear PDEs via semidefinite optimization. *Math. Program.* **108**(1), 135–158 (2006)
- Bertsimas, D., Cory-Wright, R.: On polyhedral and second-order cone decompositions of semidefinite optimization problems. *Oper. Res. Lett.* **48**(1), 78–85 (2020)
- Bhatt, A.G., Borkar, V.S.: Occupation measures for controlled markov processes: characterization and optimality. *Ann. Probab.* **24**(3), 1531–1562 (1996)
- Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
- Braun, G., Fiorini, S., Pokutta, S., Steurer, D.: Approximation limits of linear programs (beyond hierarchies). *Math. Oper. Res.* **40**(3), 756–772 (2015)
- Butcher, J.C.: *The Numerical Analysis of Ordinary Differential Equations: Runge–Kutta and General Linear Methods*. Wiley-Interscience, New York (1987)
- Cibulka, V., Korda, M., Haniš, T.: Spatio-temporal decomposition of sum-of-squares programs for the region of attraction and reachability. *IEEE Control Syst. Lett.* **6**, 812–817 (2021)
- Cuthrell, J.E., Biegler, L.T.: On the Optimization of Differential-Algebraic Process Systems. *AIChE J.* **33**(8), 1257–1270 (1987)
- Del Vecchio, D., Dy, A.J., Qian, Y.: Control theory meets synthetic biology. *J. R. Soc. Interface* **13**(120), 20160380 (2016)

20. Dowdy, G.R.: Using semidefinite programming to bound distributions in chemical engineering systems. Ph.D. thesis, Massachusetts Institute of Technology (2019)
21. Dowdy, G.R., Barton, P.I.: Bounds on stochastic chemical kinetic systems at steady state. *J. Chem. Phys.* **148**(8), 84106 (2018)
22. Dowdy, G.R., Barton, P.I.: Dynamic bounds on stochastic chemical kinetic systems using semidefinite programming. *J. Chem. Phys.* **149**(7), 74103 (2018)
23. Dunning, I., Huchette, J., Lubin, M.: JuMP: A Modeling Language for Mathematical Optimization. *SIAM Rev.* **59**(2), 295–320 (2017)
24. Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S.: Stochastic gene expression in a single cell. *Science* **297**(5584), 1183–1186 (2002)
25. Ghusinga, K.R., Vargas-Garcia, C.A., Lamperski, A., Singh, A.: Exact lower and upper bounds on stationary moments in stochastic biochemical systems. *Phys. Biol.* **14**(4), 04LT01 (2017)
26. Gillespie, C.S.: Moment-closure approximations for mass-action models. *IET Syst. Biol.* **3**(1), 52–58 (2009)
27. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**(4), 403–434 (1976)
28. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25), 2340–2361 (1977)
29. Gillespie, D.T.: A rigorous derivation of the chemical master equation. *Physica A* **188**(1–3), 404–425 (1992)
30. Gillespie, D.T., Hellander, A., Petzold, L.R.: Perspective: stochastic algorithms for chemical kinetics. *J. Chem. Phys.* **138**(17), 170901 (2013)
31. Grima, R.: A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.* **136**(15), 154105 (2012)
32. Helmes, K., Röhl, S.: A geometrical characterization of multidimensional Hausdorff polytopes with applications to exit time problems. *Math. Oper. Res.* **33**(2), 315–326 (2008)
33. Helmes, K., Röhl, S., Stockbridge, R.H.: Computing moments of the exit time distribution for Markov processes by linear programming. *Oper. Res.* **49**(4), 516–530 (2001)
34. Henrion, D.: Semidefinite characterisation of invariant measures for one-dimensional discrete dynamical systems. *Kybernetika* **48**(6), 1089–1099 (2012)
35. Henrion, D., Kirschner, F., De Klerk, E., Korda, M., Lasserre, J.B., Magron, V.: Revisiting semidefinite programming approaches to options pricing: complexity and computational perspectives. *INFORMS J. Comput.* **35**(2), 335–349 (2023)
36. Henrion, D., Korda, M.: Convex computation of the region of attraction of polynomial control systems. *IEEE Trans. Autom. Control* **59**(2), 297–312 (2014)
37. Henrion, D., Korda, M., Kružík, M., Rios-Zertuche, R.: Occupation measure relaxations in variational problems: the role of convexity. arXiv preprint [arXiv:2303.02434](https://arxiv.org/abs/2303.02434) (2023)
38. Henrion, D., Korda, M., Lasserre, J.B.: *The Moment-SOS Hierarchy: Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs*, vol. 4. World Scientific, Singapore (2020)
39. Kashima, K., Kawai, R.: Polynomial programming approach to weak approximation of Lévy-driven stochastic differential equations with application to option pricing. In: 2009 ICCAS-SICE, pp. 3902–3907. IEEE (2009)
40. Keeling, M.J.: Multiplicative moments and measures of persistence in ecology. *J. Theor. Biol.* **205**(2), 269–281 (2000)
41. Korda, M., Henrion, D., Lasserre, J.B.: Moments and convex optimization for analysis and control of nonlinear PDEs. In: *Handbook of Numerical Analysis*, vol. 23, pp. 339–366. Elsevier (2022)
42. Kuntz, J., Ottobre, M., Stan, G.B., Barahona, M.: Bounding stationary averages of polynomial diffusions via semidefinite programming. *SIAM J. Sci. Comput.* **38**(6), A3891–A3920 (2016)
43. Kuntz, J., Thomas, P., Stan, G.B., Barahona, M.: Bounding the stationary distributions of the chemical master equation via mathematical programming. *J. Chem. Phys.* **151**(3), 34109 (2019)
44. Lamperski, A., Dhople, S.: A semidefinite programming method for moment approximation in stochastic differential algebraic systems. In: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), pp. 2455–2460. IEEE (2017)
45. Lasserre, J.B.: Global Optimization with Polynomials and the Problem of Moments. *SIAM J. Optim.* **11**(3), 796–817 (2001)
46. Lasserre, J.B.: *Moments, Positive Polynomials and Their Applications*, vol. 1. World Scientific, Singapore (2010)

47. Lasserre, J.B., Henrion, D., Prieur, C., Trélat, E.: Nonlinear optimal control via occupation measures and LMI-relaxations. *SIAM J. Control. Optim.* **47**(4), 1643–1666 (2008)
48. Lasserre, J.B., Prieto-Rumeau, T., Zervos, M.: Pricing a class of exotic options via moments and SDP relaxations. *Math. Financ.* **16**(3), 469–494 (2006)
49. Liu, Q., Jia, Y.: Fluctuations-induced switch in the gene transcriptional regulatory system. *Phys. Rev. E* **70**(4), 41907 (2004)
50. Munsky, B., Khammash, M.: The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* **124**(4), 44104 (2006)
51. Na, S., Shin, S., Anitescu, M., Zavala, V.M.: On the convergence of overlapping schwarz decomposition for nonlinear optimal control. *IEEE Trans. Autom. Control* **67**(11), 5996–6011 (2022)
52. Nåsell, I.: An extension of the moment closure method. *Theor. Popul. Biol.* **64**(2), 233–239 (2003)
53. Nesterov, Y.: Squared functional systems and optimization problems. In: Frenk, H., Roos, K., Terlaky, T., Zhang, S. (eds.) *High performance optimization*, pp. 405–440. Springer, New York (2000)
54. Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology (2000)
55. Peralta-Yahya, P.P., Zhang, F., Del Cardayre, S.B., Keasling, J.D.: Microbial engineering for the production of advanced biofuels. *Nature* **488**(7411), 320–328 (2012)
56. Powers, V., Reznick, B.: Polynomials that are positive on an interval. *Trans. Am. Math. Soc.* **352**(10), 4677–4692 (2000)
57. Resnick, S.I.: *Adventures in Stochastic Processes*. Springer, New York (1992)
58. Rudin, W.: *Principles of Mathematical Analysis*, vol. 3. McGraw-Hill, New York (1976)
59. Sakurai, Y., Hori, Y.: A convex approach to steady state moment analysis for stochastic chemical reactions. In: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), pp. 1206–1211. IEEE (2017)
60. Sakurai, Y., Hori, Y.: Optimization-based synthesis of stochastic biocircuits with statistical specifications. *J. R. Soc. Interface* **15**(138), 20170709 (2018)
61. Sakurai, Y., Hori, Y.: Bounding transient moments of stochastic chemical reactions. *IEEE Control Syst. Lett.* **3**(2), 290–295 (2019)
62. Sakurai, Y., Hori, Y.: Interval analysis of worst-case stationary moments for stochastic chemical reactions with uncertain parameters. *Automatica* **146**, 110647 (2022)
63. Schnoerr, D., Sanguinetti, G., Grima, R.: Validity conditions for moment closure approximations in stochastic chemical kinetics. *J. Chem. Phys.* **141**(8), 084103 (2014)
64. Schnoerr, D., Sanguinetti, G., Grima, R.: Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.* **143**(18), 185101 (2015)
65. Shin, S., Zavala, V.M., Anitescu, M.: Decentralized schemes with overlap for solving graph-structured optimization problems. *IEEE Trans. Control Netw. Syst.* **7**(3), 1225–1236 (2020)
66. Sinha, J., Reyes, S.J., Gallivan, J.P.: Reprogramming bacteria to seek and destroy an herbicide. *Nat. Chem. Biol.* **6**(6), 464 (2010)
67. Smadbeck, P., Kaznessis, Y.N.: A closure scheme for chemical master equations. *Proc. Natl. Acad. Sci.* **110**(35), 14261–14265 (2013)
68. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **11**(1–4), 625–653 (1999)
69. Toh, K.C., Todd, M.J., Tütüncü, R.H.: SDPT3-a MATLAB software package for semidefinite programming, version 1.3. *Optim. Methods Softw.* **11**(1–4), 545–581 (1999)