MODELING AND ANALYSIS OF STOCHASTIC HYBRID SYSTEMS

João P. Hespanha

Center for Control, Dynamical Systems, and Computation
University of California, Santa Barbara, CA 93101

hespanha@ece.ucsb.edu
http://www.ece.ucsb.edu/~hespanha

Abstract

The author describes a model for Stochastic Hybrid Systems (SHSs) where transitions between discrete modes are triggered by stochastic events. The rate at which these transitions occur is allowed to depend both on the continuous and the discrete states of the SHS. Several examples of SHSs arising from a varied pool of application areas are discussed. These include modeling of the Transmission Control Protocol's (TCP) algorithm for congestion control both for long-lived and on-off flows; state-estimation for networked control systems; and the stochastic modeling of chemical reactions. These examples illustrate the use of SHSs as a modeling tool.

Attention is mostly focused on polynomial stochastic hybrid systems (pSHSs) that generally correspond to stochastic hybrid systems with polynomial continuous vector fields, reset maps, and transition intensities. For pSHSs, the dynamics of the statistical moments of the continuous states evolve according to infinite-dimensional linear ordinary differential equations (ODEs). We show that these ODEs can be approximated by finite-dimensional nonlinear ODEs with arbitrary precision. Based on this result, a procedure to build this type of approximations for certain classes of pSHSs is provided. This procedure is applied to several examples and the accuracy of the results obtained is evaluated through comparisons with Monte Carlo simulations.

I. INTRODUCTION

Hybrid systems are characterized by a state-space that can be partitioned into a continuous domain (typically $\mathbb{R}^n$) and a discrete set (typically finite). For the stochastic hybrid systems considered here, both the continuous and the discrete components of the state are stochastic processes. The evolution of the continuous-state is determined by a stochastic differential equation and the evolution of the discrete-state by a transition or reset map. The discrete transitions are triggered by stochastic events much like transitions between states of a continuous-time Markov chains. However, the rate at which these transitions occur may depend on the continuous-state. The model used

This version of the paper differs from the one that appeared in print in that we corrected a typo in Example 3. The author would like to thank Shaunak Bopardikar for making us aware of it.

This material is based upon work supported by the National Science Foundation under grants CCR-0311084, ANI-0322476.

here for SHSs, whose formal definition can be found in Sec. II, was introduced in [1] and is heavily inspired by the Piecewise-Deterministic Markov Processes (PDPs) in [2]. Alternative models can be found in [3–5].

The extended generator of a stochastic hybrid system allows one to compute the time-derivative of a "test function" of the state of the SHS along solutions to the system, and can be viewed as a generalization of the Lie derivative for deterministic systems [1, 2]. Polynomial stochastic hybrid systems (pSHSs) are characterized by extended generators that map polynomial test functions into polynomials. This happens, e.g., when the continuous vector fields, the reset maps, and the transition intensities are all polynomial functions of the continuous state. An important property of pSHSs is that if one creates an infinite vector containing the probabilities of all discrete modes, as well as all the multi-variable statistical moments of the continuous state, the dynamics of this vector are governed by an infinite-dimensional linear ordinary differential equation (ODE), which we call the *infinite-dimensional moment dynamics* and define formally in Sec. IV.

SHSs can model large classes of systems but their formal analysis presents significant challenges. Although it is straightforward to write partial differential equations (PDEs) that express the evolution of the probability distribution function for their states, generally these PDEs do not admit analytical solutions. The infinite-dimensional moment dynamics provides an alternative characterization for the distribution of the state of a pSHS. Although generally statistical moments do not provide a description of a stochastic process as accurate as the probability distribution, results such as Tchebycheff, Markoff, or Bienaymé inequalities [6, pp. 114–115] can be used to infer properties of the distribution from its moments.

In general, the infinite-dimensional *linear* ODEs that describe the moment dynamics for pSHSs are still not easy to solve analytically. However, sometimes they can be accurately approximated by a finite-dimensional *nonlinear* ODE, which we call the *truncated moment dynamics.* We show in Sec. IV that, under suitable stability assumptions, it is in principle possible for a finite-dimensional nonlinear ODE to approximate the infinite-dimensional moment dynamics, up to an error that can be made arbitrarily small. This result is based on a Taylor series expansion that is used to prove that the difference between the solutions to the finite- and infinite-dimensional ODEs can be made arbitrarily small on a compact time interval $[0, T]$. Subsequently, this is extended to the unbounded time interval $[0, \infty)$ using an argument similar to the one found in [7] to analyze two time scale systems.

Aside from its theoretical interest, the above mentioned result motivates a procedure to actually construct finite-dimensional approximations for certain classes of pSHSs. This procedure, which is described in Sec. V, is applicable to pSHSs for which the (infinite) matrix that characterizes the moment dynamics exhibits a certain diagonal-band structure and appropriate decoupling between certain moments of distinct discrete modes. The details of this structure can be found in Lemma 1.

To illustrate the applicability of the results, we consider several systems that appeared in the literature and that can be modeled by pSHSs (cf. Sec. III). For each example, we construct in Sec. VI truncated moment dynamics and evaluate how they compare with estimates for the moments obtained from a large number of Monte Carlo simulations. The examples considered include:

1) The modeling of the sending rate for network traffic under TCP congestion control. We consider two distinct cases: long-lived traffic corresponding to the transfer of files with infinite length; and on-off traffic consisting of file transfers with exponentially distributed lengths, alternated by times of inactivity (also exponentially distributed). These examples are motivated by [1, 8].
2) The modeling of the state-estimation error in a networked control system that occasionally receives state measurements over a communication network. The rate at which the measurements are transmitted depends on the current estimation error. This type of scheme was shown to out-perform periodic transmission and can actually be used to approximate an optimal transmission scheme [9, 10].
3) Gillespie's stochastic simulation algorithm (SSA) for chemical reactions [11], which describes the evolution of the number of particles involved in a set of reactions. The reactions analyzed were taken from [12, 13] and are particularly difficult to simulate due to the existence of two very distinct time scales.

A few other academic examples with tutorial value are also included. A subset of the results and examples presented here appeared in [14].

II. Stochastic Hybrid Systems Modeling

A SHS is defined by a stochastic differential equation (SDE)

$$\dot{\mathbf{x}} = f(\mathbf{q},\mathbf{x},t) + g(\mathbf{q},\mathbf{x},t)\dot{\mathbf{n}}, \qquad f : \mathcal{Q}\times\mathbb{R}^n\times[0,\infty)\to\mathbb{R}^n, \quad g : \mathcal{Q}\times\mathbb{R}^n\times[0,\infty)\to\mathbb{R}^{n\times k}, \tag{1}$$

a family of m *discrete transition/reset maps*

$$(\mathbf{q},\mathbf{x}) = \phi_\ell(\mathbf{q}^-,\mathbf{x}^-,t), \qquad \phi_\ell : \mathcal{Q}\times\mathbb{R}^n\times[0,\infty)\to\mathcal{Q}\times\mathbb{R}^n, \;\; \forall \ell\in\{1,\dots,m\}, \tag{2}$$

and a family of m *transition intensities*

$$\lambda_\ell(\mathbf{q},\mathbf{x},t), \qquad \lambda_\ell : \mathcal{Q}\times\mathbb{R}^n\times[0,\infty)\to[0,\infty), \;\; \forall \ell\in\{1,\dots,m\}, \tag{3}$$

where $\mathbf{n}$ denotes a k-vector of independent Brownian motion processes and $\mathcal{Q}$ a (typically finite) discrete set. A SHS characterizes a jump process $\mathbf{q} : [0,\infty)\to\mathcal{Q}$ called the *discrete state*; a stochastic process $\mathbf{x} : [0,\infty)\to\mathbb{R}^n$ with piecewise continuous sample paths called the *continuous state*; and m stochastic counters $\mathbf{N}_\ell : [0,\infty)\to\mathbb{N}$ called the *transition counters*. In this paper, $\mathbb{R}$ denotes the set of real numbers and $\mathbb{N}$ the set of nonnegative integers.

In essence, between transition counter increments the discrete state remains constant whereas the continuous state flows according to (1). At transition times, the continuous and discrete states are reset according to (2). Each transition counter $\mathbf{N}_\ell$ counts the number of times that the corresponding discrete transition/reset map ϕ_ℓ is "activated." The frequency at which this occurs is determined by the transition intensities (3). In particular, the probability that the counter $\mathbf{N}_\ell$ will increment in an "elementary interval" $(t, t+dt]$, and therefore that the corresponding transition takes place, is given by $\lambda_\ell(\mathbf{q}(t),\mathbf{x}(t),t)dt$. In practice, one can think of the intensity of a transition as the instantaneous rate at which that transition occurs. The reader is referred to [1] for a mathematically precise characterization of this SHS.

It is often convenient to represent SHSs by a directed graph as in Figure 1, where each node corresponds to a discrete mode and each edge to a transition between discrete modes. The nodes are labeled with the corresponding discrete mode and the vector fields that determines the evolution of the continuous state in that particular mode. The start of each edge is labeled with the probability of transition in an elementary interval and the end of each edge is labeled with the corresponding reset-map.

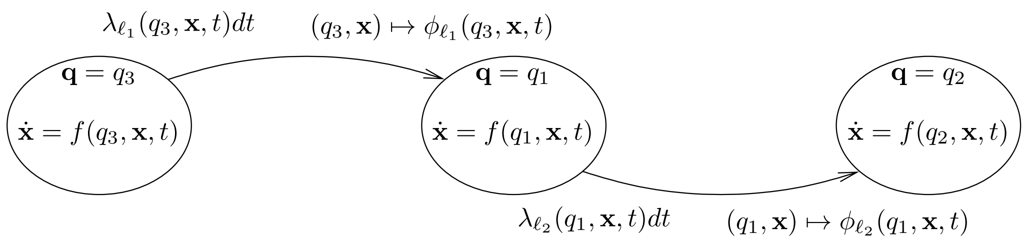


Fig. 1. Graphical representation of a stochastic hybrid system

The following result can be used to compute expectations on the state of a SHS. For briefness, we omit a few technical assumptions that are straightforward to verify for the SHSs considered here:

Theorem 1 ([1]): Given a function $\psi : \mathcal{Q} \times \mathbb{R}^n \times [0, \infty) \to \mathbb{R}$ that is twice continuously differentiable with respect to its second argument and once continuously differentiable with respect to the third one, we have that

$$\frac{\mathrm{d}\,\mathrm{E}[\psi(\mathbf{q}(t), \mathbf{x}(t), t)]}{\mathrm{d}t} = \mathrm{E}[(L\psi)(\mathbf{q}(t), \mathbf{x}(t), t)], \tag{4}$$

where $\forall (q, x, t) \in \mathcal{Q} \times \mathbb{R}^n \times [0, \infty)$

$$\begin{aligned}(L\psi)(q,x,t) := \frac{\partial \psi(q,x,t)}{\partial x} f(q,x,t) + \frac{\partial \psi(q,x,t)}{\partial t} + \\ + \frac{1}{2}\operatorname{trace}\Big(\frac{\partial^2 \psi(q,x)}{\partial x^2} g(q,x,t) g(q,x,t)'\Big) + \\ + \sum_{\ell=1}^{m} \Big(\psi\big(\phi_\ell(q,x,t), t\big) - \psi(q,x,t)\Big)\lambda_\ell(q,x,t),\end{aligned} \tag{5}$$

and $\frac{\partial \psi(q,x,t)}{\partial t}$, $\frac{\partial \psi(q,x,t)}{\partial x}$, and $\frac{\partial^2 \psi(q,x)}{\partial x^2}$ denote the partial derivative of $\psi(q, x, t)$ with respect to t, the gradient of $\psi(q, x, t)$ with respect to x, and the Hessian matrix of ψ with respect to x, respectively. The operator $\psi \mapsto L\psi$ defined by (5) is called the *extended generator* of the SHS. □

We say that a SHS is *polynomial (pSHS)* if its extended generator L is closed on the set of finite-polynomials in x, i.e., $(L\psi)(q, x, t)$ is a finite-polynomial in x for every finite-polynomial $\psi(q, x, t)$ in x. By a *finite-polynomials in* x we mean a function $\psi(q, x, t)$ such that $x \mapsto \psi(q, x, t)$ is a (multi-variable) polynomial of finite degree for each fixed $q \in \mathcal{Q}$, $t \in [0, \infty)$. A pSHS is obtained, e.g., when the vector fields f and g, the reset maps ϕ_ℓ, and the transition intensities λ_ℓ are all finite-polynomials in x.

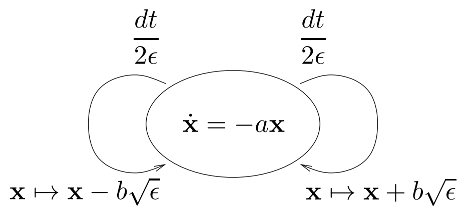


Fig. 2. Random walk SHS in Example 1

III. EXAMPLES OF POLYNOMIAL STOCHASTIC HYBRID SYSTEMS

Example 1 (Random walk): Let $\mathbf{x} \in \mathbb{R}$ denote the velocity of a particle moving in a viscous medium that every so often is instantaneously increased or decrease by a fixed amount $b\sqrt{\epsilon}$, with $b \in \mathbb{R}$, $\epsilon > 0$. The interval of time between increases (as well as between decreases) is exponentially distributed with mean $2\epsilon > 0$. The process $\mathbf{x}$ can be generated by a SHS (represented graphically in Figure 2) with continuous dynamics $\dot{\mathbf{x}} = -a\mathbf{x}$ and two reset maps $\mathbf{x} \mapsto \phi_{1,2}(\mathbf{x}) := \mathbf{x} \pm b\sqrt{\epsilon}$ both with constant transitions intensities $\lambda_{1,2}(\mathbf{x}) := \frac{1}{2\epsilon}$. The constant $a \geq 0$ accounts for viscous friction. This SHS has a single discrete mode that we omitted for simplicity and its generator is given by

$$(L\psi)(x) = -ax\frac{\partial\psi(x)}{\partial x} + \frac{\psi(x + b\sqrt{\epsilon}) + \psi(x - b\sqrt{\epsilon}) - 2\psi(x)}{2\epsilon},$$

which is closed on the set of finite-polynomials in x. It is well known that as $\epsilon \to 0$, $\mathbf{x}$ approaches the solution to the SDE $\dot{\mathbf{x}} = a\mathbf{x} + b\dot{\mathbf{n}}$, where $\mathbf{n}$ is a Brownian motion process. This SDE is easy to solve analytically, at least when the initial distribution of $\mathbf{x}$ is Gaussian. Our (somewhat academic) goal here would be to study what happens when ϵ is not necessarily close to zero and when the distribution of the initial state is arbitrary. $\square$

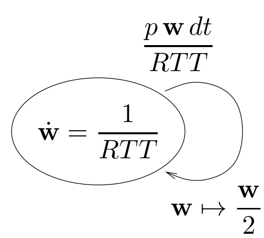


Fig. 3. TCP long-lived SHS in Example 2

Example 2 (TCP long-lived [15]): The congestion window size $\mathbf{w} \in [0, \infty)$ of a long-lived TCP flow can be generated by a SHS (represented graphically in Figure 3) with continuous dynamics $\dot{\mathbf{w}} = \frac{1}{RTT}$ and a reset map $\mathbf{w} \mapsto \frac{\mathbf{w}}{2}$, with intensity $\lambda(\mathbf{w}) := \frac{p\,\mathbf{w}}{RTT}$. The round-trip-time RTT and the drop-rate p are parameters that we assume constant. This SHS has a single discrete mode that we omit for simplicity and its generator is given by

$$(L\psi)(w) = \frac{1}{RTT}\frac{\partial\psi(w)}{\partial w} + \frac{pw\big(\psi(w/2) - \psi(w)\big)}{RTT},$$

which is closed on the set of finite-polynomials in w. $\square$

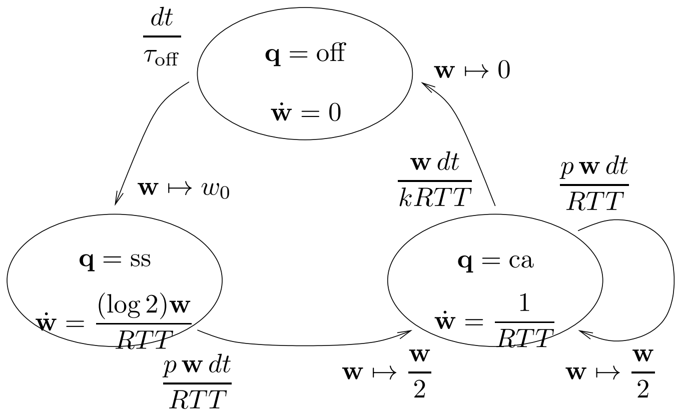


Fig. 4. TCP on-off SHS in Example 3

Example 3 (TCP on-off [15]): The congestion window size $\mathbf{w} \in [0,\infty)$ for a stream of TCP flows separated by inactivity periods can be generated by a SHS (represented graphically in Figure 4) with three discrete modes $\mathcal{Q} := \{\text{ss}, \text{ca}, \text{off}\}$, one corresponding to slow-start, another to congestion avoidance, and the final one to flow inactivity. Its continuous dynamics are defined by

$$\dot{\mathbf{w}} = \begin{cases} \frac{(\log 2)\mathbf{w}}{RTT} & \mathbf{q} = \text{ss} \\ \frac{1}{RTT} & \mathbf{q} = \text{ca} \\ 0 & \mathbf{q} = \text{off}; \end{cases}$$

the reset maps associated with packet drops, end of flows, and start of flows are given by

$$(\mathbf{q},\mathbf{w}) \mapsto \phi_1(\mathbf{q},\mathbf{w}) := \big(\text{ca}, \frac{\mathbf{w}}{2}\big), \qquad (\mathbf{q},\mathbf{w}) \mapsto \phi_2(\mathbf{q},\mathbf{w}) := (\text{off}, 0), \qquad (\mathbf{q},\mathbf{w}) \mapsto \phi_3(\mathbf{q},\mathbf{w}) := (\text{ss}, w_0)$$

respectively; and the corresponding intensities are

$$\lambda_1(\mathbf{q},\mathbf{w}) := \begin{cases} \frac{p\,\mathbf{w}}{RTT} & \mathbf{q} \in \{\text{ss},\text{ca}\} \\ 0 & \mathbf{q} = \text{off} \end{cases} \quad \lambda_2(\mathbf{q},\mathbf{w}) := \begin{cases} \frac{\mathbf{w}}{kRTT} & \mathbf{q} = \text{ca} \\ 0 & \mathbf{q} = \{\text{ss},\text{off}\} \end{cases} \quad \lambda_3(\mathbf{q},\mathbf{w}) := \begin{cases} \frac{1}{\tau_{\text{off}}} & \mathbf{q} = \text{off} \\ 0 & \mathbf{q} \in \{\text{ss},\text{ca}\}. \end{cases}$$

The round-trip-time RTT, the drop-rate p, the average file size k (exponentially distributed), the average off-time τ_{off} (also exponentially distributed), and the initial window size w_0 are parameters that we assume constant. The generator for this SHS is given by

$$(L\psi)(q,w) = \begin{cases} \frac{(\log 2)w}{RTT}\frac{\partial\psi(\text{ss},w)}{\partial w} + \frac{pw\left(\psi(\text{ca},w/2)-\psi(\text{ss},w)\right)}{RTT} + \frac{w\left(\psi(\text{off},0)-\psi(\text{ss},w)\right)}{k\,RTT} & q = \text{ss} \\ \frac{1}{RTT}\frac{\partial\psi(\text{ca},w)}{\partial w} + \frac{pw\left(\psi(\text{ca},w/2)-\psi(\text{ca},w)\right)}{RTT} & q = \text{ca} \\ \frac{\psi(\text{ss},w_0)-\psi(\text{off},w)}{\tau_{\text{off}}} & q = \text{off}, \end{cases}$$

which is closed on the set of finite-polynomials in w. □

Example 4 (Networked control system [9]): Suppose that the state of a stochastic scalar linear system $\dot{\mathbf{x}} = a\,\mathbf{x} + b\,\dot{\mathbf{n}}$ is estimated based on state-measurements received through a network. For simplicity we assume that state

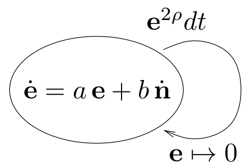


Fig. 5. Networked control system SHS in Example 4

measurements are noiseless and delay free. The corresponding state estimation error $\mathbf{e} \in \mathbb{R}$ can be generated by a SHS (represented graphically in Figure 5) with continuous dynamics $\dot{\mathbf{e}} = a\,\mathbf{e} + b\,\dot{\mathbf{n}}$ and one reset map $\mathbf{e} \mapsto 0$ that is activated whenever a state measurement is received. It was conjectured in [9] and later shown in [10] that transmitting measurements at a rate that depends on the state-estimation error is optimal when one wants to minimize the variance of the estimation error, while penalizing the average rate at which messages are transmitted. This motivates considering a reset intensity of the form $\lambda(\mathbf{e}) := \mathbf{e}^{2\rho}$, for some $\rho \in \mathbb{N}$. This SHS has a single discrete mode that we omitted for simplicity and its generator is given by

$$(L\psi)(e) := a\,e\frac{\partial\psi(e)}{\partial e} + \frac{b^2}{2}\frac{\partial^2\psi(e)}{\partial e^2} + \big(\psi(0) - \psi(e)\big)e^{2\rho},$$

which is closed on the set of finite-polynomials in e. □

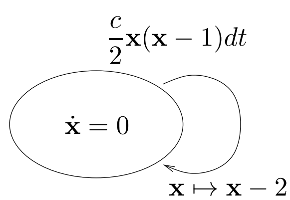


Fig. 6. Decaying reaction SHS in Example 5

Example 5 (Decaying reaction): The number of particles $\mathbf{x}$ of a species S_1 involved in a decaying reaction

$$2S_1 \xrightarrow{c} 0$$

can be generated by a SHS (represented graphically in Figure 6) with continuous dynamics $\dot{\mathbf{x}} = 0$ and reset map $\mathbf{x} \mapsto \mathbf{x} - 2$ with intensity $\lambda(\mathbf{x}) := \frac{c}{2}\mathbf{x}(\mathbf{x} - 1)$. Since the number of particles takes values in the discrete set of integers, we can regard $\mathbf{x}$ as either part of the discrete or the continuous state. We choose to regard it as a continuous variable because we are interested in studying its statistical moments. In this case, the SHS has a single discrete mode that we omit for simplicity. This model is exactly equivalent to **(author?)**'s [11] stochastic simulation algorithm (SSA) and its generator is given by

$$(L\psi)(x) = \frac{c}{2}x(x-1)\big(\psi(x-2) - \psi(x)\big),$$

which is closed on the set of finite-polynomials in x. □

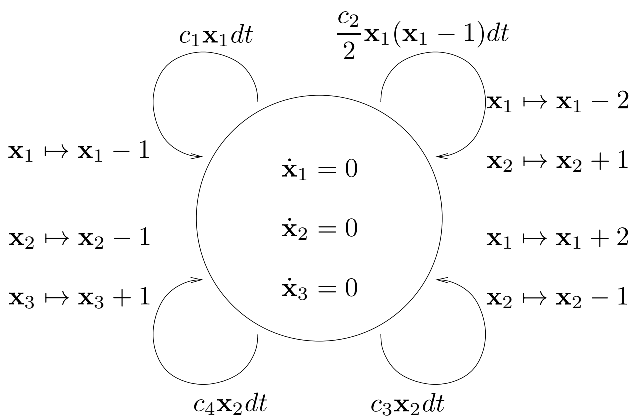


Fig. 7. Decaying reaction SHS in Example 6

Example 6 (Decaying-dimerizing reaction set [12, 13]): The number of particles $\mathbf{x} := (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ of three species involved in the following set of decaying-dimerizing reactions

$$S_1 \xrightarrow{c_1} 0, \qquad 2S_1 \xrightarrow{c_2} S_2, \qquad S_2 \xrightarrow{c_3} 2S_1, \qquad S_2 \xrightarrow{c_4} S_3 \tag{6}$$

can be generated by a SHS (represented graphically in Figure 7) with continuous dynamics $\dot{\mathbf{x}} = 0$ and four reset maps

$$\mathbf{x} \mapsto \phi_1(\mathbf{x}) := \begin{bmatrix} \mathbf{x}_1 - 1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} \quad \mathbf{x} \mapsto \phi_2(\mathbf{x}) := \begin{bmatrix} \mathbf{x}_1 - 2 \\ \mathbf{x}_2 + 1 \\ \mathbf{x}_3 \end{bmatrix} \qquad \mathbf{x} \mapsto \phi_3(\mathbf{x}) := \begin{bmatrix} \mathbf{x}_1 + 2 \\ \mathbf{x}_2 - 1 \\ \mathbf{x}_3 \end{bmatrix} \quad \mathbf{x} \mapsto \phi_4(\mathbf{x}) := \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 - 1 \\ \mathbf{x}_3 + 1 \end{bmatrix}$$

with intensities

$$\lambda_1(\mathbf{x}) := c_1\mathbf{x}_1, \qquad \lambda_2(\mathbf{x}) := \frac{c_2}{2}\mathbf{x}_1(\mathbf{x}_1 - 1), \qquad \lambda_3(\mathbf{x}) := c_3\mathbf{x}_2, \qquad \lambda_4(\mathbf{x}) := c_4\mathbf{x}_2,$$

respectively. Also here we choose to regard the $\mathbf{x}_i$ as continuous variables because we are interested in studying their statistical moments. The generator for this SHS is given by

$$\begin{aligned}(L\psi)(x) = c_1x_1\big(\psi(x_1-1,x_2,x_3) - \psi(x)\big) + \frac{c_2}{2}x_1(x_1-1)\big(\psi(x_1-2,x_2+1,x_3) - \psi(x)\big) \\ + c_3x_2\big(\psi(x_1+2,x_2-1,x_3) - \psi(x)\big) + c_4x_2\big(\psi(x_1,x_2-1,x_3+1) - \psi(x)\big),\end{aligned}$$

which is closed on the set of finite-polynomials in x. □

IV. Moment Dynamics

To fully characterize the dynamics of a SHS one would like to determine the evolution of the probability distribution for its state $(\mathbf{q}, \mathbf{x})$. In general, this is difficult so a more reasonable goal is to determine the evolution of (i) the probability of $\mathbf{q}(t)$ being on each mode and (ii) the moments of $\mathbf{x}(t)$ conditioned to $\mathbf{q}(t)$. In fact, often one can even get away with only determining a few low-order moments and then using results such as Tchebycheff, Markoff, or Bienaymé inequalities [6, pp. 114–115] to infer properties of the overall distribution.

Given a discrete state $\bar{q} \in \mathcal{Q}$ and a vector of n integers $m = (m_1, m_2, \dots, m_n) \in \mathbb{N}^n$, we define the *test-function associated with $\bar{q}$ and m* to be

$$\psi_{\bar{q}}^{(m)}(q,x) := \begin{cases} x^{(m)} & q = \bar{q} \\ 0 & q \neq \bar{q}, \end{cases} \qquad \forall q \in \mathcal{Q}, x \in \mathbb{R}^n$$

and the *(uncentered) moment associated with $\bar{q}$ and m* to be

$$\mu_{\bar{q}}^{(m)}(t) := \mathrm{E}\left[\psi_{\bar{q}}^{(m)}\big(\mathbf{q}(t), \mathbf{x}(t)\big)\right] \qquad \forall t \geq 0. \tag{7}$$

Here and in the sequel, given a vector $x = (x_1, x_2, \dots, x_n)$, we use $x^{(m)}$ to denote the monomial $x_1^{m_1} x_2^{m_2} \cdots x_n^{m_n}$.

PSHSs have the property that if one stacks all moments in an infinite vector μ_∞, its dynamics can be written as

$$\dot{\mu}_\infty = A_\infty(t)\mu_\infty \qquad \forall t \geq 0, \tag{8}$$

for some appropriately defined infinite matrix $A_\infty(t)$. This is because $\forall \bar{q} \in \mathcal{Q}, m = (m_1, \dots, m_n) \in \mathbb{N}^n$, the expression $(L\psi_{\bar{q}}^{(m)})(q,x,t)$ is a finite-polynomial in x and therefore can be written as a finite linear combination of test-functions (possibly with time-varying coefficients). Taking expectations on this linear combination and using (4), (7), we conclude that $\dot{\mu}_{\bar{q}}^{(m)}$ can be written as a linear combination of uncentered moments in μ_∞, leading to (8). In the sequel, we refer to (8) as the *infinite-dimensional moment dynamics*. Analyzing (and even simulating) (8) is generally difficult. However, as mentioned above one can often get away with just computing a few low-order moments. One would therefore like to determine a finite-dimensional system of ODEs that describes the evolution of a few low-order models, perhaps only approximately.

When the matrix A_∞ is lower triangular (e.g., as in Example 1 and Example 4 with $\rho = 0$), one can simply truncate the vector μ_∞ by dropping all but its first k elements and obtain a finite-dimensional system that exactly describes the evolution of the moments[1]. However, in general A_∞ has nonzero elements above the main diagonal and therefore if one defines $\mu \in \mathbb{R}^k$ to contain the top k elements of μ_∞, one obtains from (8) that

$$\dot{\mu} = I_{k\times\infty} A_\infty(t)\mu_\infty = A(t)\mu + B(t)\bar{\mu}, \qquad \bar{\mu} = C\mu_\infty, \tag{9}$$

where $I_{k\times\infty}$ denotes a matrix composed of the first k rows of the infinite identity matrix, $\bar{\mu} \in \mathbb{R}^r$ contains all the moments that appear in the first k elements of $A_\infty(t)\mu_\infty$ but that do not appear in μ, and C is the projection matrix that extracts $\bar{\mu}$ from μ_∞. Our goal is to approximate the infinite dimensional system (8) by a finite-dimensional nonlinear ODE of the form

$$\dot{\nu} = A(t)\nu + B(t)\bar{\nu}(t), \quad \bar{\nu} = \varphi(\nu, t), \tag{10}$$

[1] Other truncations are possible, but it is often convenient to keep the low-order moments of $\mathbf{x}$ as the states of the truncated system because these are physically meaningful.

where the map $\varphi : \mathbb{R}^k \times [0,\infty) \to \mathbb{R}^r$ should be chosen so as to keep $\nu(t)$ close to $\mu(t)$. We call (10) the *truncated moment dynamics* and φ the *truncation function*. We need the following two stability assumptions to establish sufficient conditions for the approximation to be valid.

Assumption 1 (Boundedness): There exist sets Ω_μ and Ω_ν such that all solutions to (8) and (10) starting at some time $t_0 \geq 0$ in Ω_μ and Ω_ν, respectively, exist and are smooth on $[t_0,\infty)$ with all derivatives of their first k elements uniformly bounded. The set Ω_ν is assumed to be forward invariant. □

Assumption 2 (Incremental Stability): There exists a function[2] $\beta \in \mathcal{KL}$ such that, for every solution μ_∞ to (8) starting in Ω_μ at some time $t_0 \geq 0$, and every $t_1 \geq t_0$, $\nu_1 \in \Omega_\nu$ there exists some $\hat{\mu}_\infty(t_1) \in \Omega_\mu$ whose first k elements match ν_1 and

$$\|\mu(t) - \hat{\mu}(t)\| \leq \beta(\|\mu(t_1) - \hat{\mu}(t_1)\|, t - t_1), \qquad \forall t \geq t_1, \tag{11}$$

where $\mu(t)$ and $\hat{\mu}(t)$ denote the first k elements of the solutions to (8) starting at $\mu_\infty(t_1)$ and $\hat{\mu}_\infty(t_1)$, respectively. □

Remark 1: Assumption 2 is essentially a requirement of uniform stability for (8), but it was purposely formulated without requiring Ω_μ to be a subset of a normed space to avoid having to choose a norm under which the (infinite) vectors of moments are bounded. It is important to emphasize that a certain "richness" of Ω_μ is implicit in this assumption. Indeed, for the existence of $\hat{\mu}_\infty(t_1) \in \Omega_\mu$ whose first components match an arbitrary $\nu_1 \in \Omega_\nu$ one needs Ω_ν to be contained in the projection of Ω_μ into its first k components. □

The result that follows establishes that the difference between solutions to (8) and (10) converges to an arbitrarily small ball, provided that a sufficiently large but *finite* number of derivatives of these signals match point-wise. To state this result, the following notation is needed: We define the matrices $C^i(t)$, $i \in \mathbb{N}$ recursively by

$$C^0(t) = C, \qquad C^{i+1}(t) = C^i(t)A_\infty(t) + \dot{C}^i(t), \qquad \forall t \geq 0,\ i \in \mathbb{N},$$

and the family of functions $L^i\varphi : \mathbb{R}^k \times [0,\infty) \to \mathbb{R}^r$, $i \in \mathbb{N}$ recursively by

$$(L^0\varphi)(\nu,t) = \varphi(\nu,t), \qquad (L^{i+1}\varphi)(\nu,t) = \frac{\partial(L^i\varphi)(\nu,t)}{\partial\nu}\big(A(t)\nu + B(t)\varphi(\nu,t)\big) + \frac{\partial(L^i\varphi)(\nu,t)}{\partial t},$$

$\forall t \geq 0,\ \nu \in \mathbb{R}^k,\ i \in \mathbb{N}$. These definitions allow us to compute time derivatives of $\bar{\mu}(t)$ and $\bar{\nu}(t)$ along solutions to (8) and (10), respectively, because

$$\frac{\mathrm{d}^i\bar{\mu}(t)}{\mathrm{d}t^i} = C^i(t)\mu_\infty(t), \qquad \frac{\mathrm{d}^i\bar{\nu}(t)}{\mathrm{d}t^i} = (L^i\varphi)(\nu(t),t), \qquad \forall t \geq 0,\ i \in \mathbb{N}. \tag{12}$$

Theorem 2 (cf. Appendix): For every $\delta > 0$, there exists an integer N sufficiently large for which the following result holds: Assuming that for every $\tau \geq 0$, $\mu_\infty \in \Omega_\mu$

$$C^i(\tau)\mu_\infty = (L^i\varphi)(\mu,\tau), \qquad \forall i \in \{0,1,\ldots,N\}, \tag{13}$$

[2]A function $\beta : [0,\infty) \times [0,\infty) \to [0,\infty)$ is of *class* $\mathcal{KL}$ if $\beta(0,t) = 0$, $\forall t \geq 0$; $\beta(s,t)$ is continuous and strictly increasing on s, $\forall t \geq 0$; and $\lim_{t\to\infty} \beta(s,t) = 0$, $\forall s \geq 0$.

where μ denotes the first k elements of μ_∞, then

$$\|\mu(t)-\nu(t)\| \leq \beta(\|\mu(t_0)-\nu(t_0)\|, t-t_0)+\delta, \qquad \forall t \geq t_0 \geq 0, \tag{14}$$

along all solutions to (8) and (10) with initial conditions $\mu_\infty(t_0) \in \Omega_\mu$ and $\nu(t_0) \in \Omega_\nu$, respectively, where $\mu(t)$ denotes the first k elements of $\mu_\infty(t)$. □

The proof of this theorem is technical and therefore we leave it for the appendix. The idea behind it is that the conditions (13) guarantees that when the solution μ_∞ to the infinite-dimensional system (8) starts in Ω_μ, the $N+1$ derivatives of its first k elements μ match those of the solution ν to the truncated system (10). Using a standard argument based on a Taylor series expansion, we thus conclude that these two solutions will remain close in a bounded interval of length T. To extend this to the unbounded interval $[0,\infty)$, we use the stability of the infinite-dimensional (8): After the first interval $[0,T]$, we can start another solution $\hat{\mu}_\infty$ to (8) on Ω_μ, whose first k elements $\hat{\mu}$ match ν at time T, and conclude from (13) that $\hat{\mu}$ and ν will remain close on $[T,2T]$. But because of stability μ and $\hat{\mu}$ will converge to each other and therefore μ and ν also remain close. This argument can be repeated to obtain the bound (14) on the unbounded time interval $[0,\infty)$. A similar argument was used in the proof of [7, Theorem 1] to analyze two time scale systems. It is important to emphasize that for this argument to hold, Ω_μ need not be forward invariant because we only use (13) for the solutions $\hat{\mu}$ that are always "re-started" in Ω_μ. This explains why one may get good matches between the original and the truncated systems, even if (13) only holds for fairly small subsets Ω_μ of the overall infinite-dimensional state-space.

V. Construction of Approximate Truncations

Given a constant $\delta > 0$ and sets Ω_μ, Ω_ν, it may be very difficult to determine the integer N for which the approximation bound (14) holds. This is because, although the proof of Theorem 2 is constructive, the computation of N requires explicit knowledge of the function $\beta \in \mathcal{KL}$ in Assumption 2 and, at least for most of the examples considered here, this assumption is difficult to verify. Nevertheless, Theorem 2 is still useful because it provides the explicit conditions (13) that the truncation function φ should satisfy for the solution to the truncated system to approximate the one of the original system. For the problems considered here we require (13) to hold for $N=1$, which corresponds to a second Taylor expansions of the solution. Note that for $N=1$, (13) simply becomes

$$C\mu_\infty = \varphi(\mu,\tau), \qquad\qquad CA_\infty(\tau)\mu_\infty = \frac{\partial\varphi(\mu,\tau)}{\partial\mu} I_{k\times\infty} A_\infty(\tau)\mu_\infty + \frac{\partial\varphi(\mu,\tau)}{\partial t}, \tag{15}$$

$\forall \mu_\infty \in \Omega_\mu,\ \tau \geq 0$. Lacking knowledge of β, we will not be able to explicitly compute for which values of δ (14) will hold, but we will show by simulation that the truncation obtained provides a very accurate approximation to the infinite-dimensional system (8), even for such a small choice of N. We restrict our attention to functions φ and sets Ω_μ for which it is simple to use (15) to explicitly compute truncated systems.

a) Separable truncation functions:: For all the examples considered, we consider functions φ of the form

$$\varphi(\nu,t)=\Lambda\nu^{(\Gamma)}:=\Lambda\begin{bmatrix}\nu_1^{\gamma_{11}}\nu_2^{\gamma_{12}}\cdots\nu_k^{\gamma_{1k}}\\ \nu_1^{\gamma_{21}}\nu_2^{\gamma_{22}}\cdots\nu_k^{\gamma_{2k}}\\ \vdots\\ \nu_1^{\gamma_{r1}}\nu_2^{\gamma_{r2}}\cdots\nu_k^{\gamma_{rk}}\end{bmatrix}, \tag{16}$$

for appropriately chosen constant matrices $\Gamma:=[\gamma_{ij}]\in\mathbb{R}^{r\times k}$ and $\Lambda\in\mathbb{R}^{r\times r}$, with Λ diagonal. In this case,

$$\begin{aligned}\frac{\partial\varphi(\nu)}{\partial\nu}&=\Lambda\begin{bmatrix}\nu_1^{\gamma_{11}}\nu_2^{\gamma_{12}}\cdots\nu_k^{\gamma_{1k}}\gamma_{11}\nu_1^{-1} & \nu_1^{\gamma_{11}}\nu_2^{\gamma_{12}}\cdots\nu_k^{\gamma_{1k}}\gamma_{12}\nu_2^{-1} & \cdots & \nu_1^{\gamma_{11}}\nu_2^{\gamma_{12}}\cdots\nu_k^{\gamma_{1k}}\gamma_{1k}\nu_k^{-1}\\ \nu_1^{\gamma_{21}}\nu_2^{\gamma_{22}}\cdots\nu_k^{\gamma_{2k}}\gamma_{21}\nu_1^{-1} & \nu_1^{\gamma_{21}}\nu_2^{\gamma_{22}}\cdots\nu_k^{\gamma_{2k}}\gamma_{22}\nu_2^{-1} & \cdots & \nu_1^{\gamma_{21}}\nu_2^{\gamma_{22}}\cdots\nu_k^{\gamma_{2k}}\gamma_{2k}\nu_k^{-1}\\ \vdots & \vdots & \ddots & \vdots\\ \nu_1^{\gamma_{r1}}\nu_2^{\gamma_{r2}}\cdots\nu_k^{\gamma_{rk}}\gamma_{r1}\nu_1^{-1} & \nu_1^{\gamma_{r1}}\nu_2^{\gamma_{r2}}\cdots\nu_k^{\gamma_{rk}}\gamma_{r2}\nu_2^{-1} & \cdots & \nu_1^{\gamma_{r1}}\nu_2^{\gamma_{r2}}\cdots\nu_k^{\gamma_{rk}}\gamma_{rk}\nu_k^{-1}\end{bmatrix}\\ &=\Lambda\,\mathrm{diag}[\nu^{(\Gamma)}]\Gamma\,\mathrm{diag}[\nu_1^{-1},\nu_2^{-1},\dots,\nu_k^{-1}]\end{aligned}$$

and therefore (15) becomes

$$C\mu_\infty=\Lambda\,\mu^{(\Gamma)}, \tag{17a}$$

$$CA_\infty(\tau)\mu_\infty=\Lambda\,\mathrm{diag}[C\mu_\infty]\Gamma\,\mathrm{diag}[\mu_1^{-1},\mu_2^{-1},\dots,\mu_k^{-1}]I_{k\times\infty}A_\infty(\tau)\mu_\infty. \tag{17b}$$

This particular form for φ is convenient because (17a) often uniquely defines Λ and then (17b) provides a linear system of equations on Γ, for which it is straightforward to determine if a solution exists.

b) Deterministic distributions:: A set Ω_μ that is particularly tractable corresponds to deterministic distributions $\mathcal{F}_{\det}:=\{P(\cdot\,;q,x):x\in\Omega_x,q\in\mathcal{Q}\}$, where $P(\cdot\,;q,x)$ denotes the distribution of $(\mathbf{q},\mathbf{x})$ for which $\mathbf{q}=q$ and $\mathbf{x}=x$ with probability one; and Ω_x a subset of the continuous state space $\mathbb{R}^n$. For a particular distribution $P(\cdot\,;q,x)$, the (uncentered) moment associated with $\bar q$ and $m\in\mathbb{N}^n$ is given by

$$\mu_{\bar q}^{(m)}:=\int\psi_{\bar q}^{(m)}(\tilde q,\tilde x)P(d\tilde q\,d\tilde x;q,x)=\psi_{\bar q}^{(m)}(q,x):=\begin{cases}x^{(m)} & q=\bar q\\ 0 & q\neq\bar q,\end{cases}$$

and therefore the vectors μ_∞ in Ω_μ have this form. Although this family of distributions may seem very restrictive, it will provide us with truncations that are accurate even when the pSHSs evolve towards very "nondeterministic" distributions, i.e., with significant variance. For this set Ω_μ, (17) takes a particularly simple form and the following result provides a simple set of conditions to test if a truncation is possible.

Lemma 1 (cf. Appendix): Let Ω_μ be the set of deterministic distributions $\mathcal{F}_{\det}$ with Ω_x containing some open ball in $\mathbb{R}^n$ and consider truncation functions φ of the form (16). The matrices Λ and Γ must satisfy:

1) Λ must be the identity matrix.
2) For every moment $\mu_{q_\ell}^{(m_\ell)}$ in $\bar\mu$ and every moment $\mu_{q_i}^{(m_i)}$ in μ for which $q_i\neq q_\ell$, one must have $\gamma_{\ell,i}=0$.
3) For every moment $\mu_{q_\ell}^{(m_\ell)}$ in $\bar\mu$, one must have $\sum_{\substack{i=1\\ q_i=q_\ell}}^{k}\gamma_{\ell,i}m_i=m_\ell$.

Moreover, the following conditions are necessary for the existence of a function φ of the form (16) that satisfies (15):

4) For every moment $\mu_{q_\ell}^{(m_\ell)}$ in $\bar{\mu}$ the polynomial[3] $\sum_{\substack{i=1\\ q_i=q_\ell}}^{\infty} a_{\ell,i}\, x^{(m_i)}$ must belong to the linear subspace generated by the polynomials

$$\Big\{ \sum_{\substack{i=1\\ q_i=q_\ell}}^{\infty} a_{j,i}\, x^{(m_\ell - m_j + m_i)} : 1 \le j \le k,\ q_j = q_\ell \Big\}.$$

5) For every moment $\mu_{q_\ell}^{(m_\ell)}$ in $\bar{\mu}$ and every moment $\mu_{q_i}^{(m_i)}$ in μ_∞ with $q_i \neq q_\ell$, we must have $a_{\ell,i} = 0$. Here we are denoting by $a_{j,i}$ the jth row, ith column entry of A_∞. □

Condition 4 imposes a diagonal-band-like structure on the submatrices of A_∞ consisting of the rows/columns that correspond to each moment that appears in $\bar{\mu}$. This condition holds for Examples 2, 3, and 4. It does not quite hold for Examples 5 and 6 but their moment dynamics can be simplified so as to satisfy this condition without introducing a significant error.

Condition 5 imposes a form of decoupling between different modes in the equations for $\dot{\bar{\mu}}$. This condition holds trivially for all examples that have a single discrete mode. It does not hold for Example 3, but also here it is possible to simplify the moment dynamics to satisfy this condition without introducing a significant error.

VI. Examples of Truncations

We now present truncated systems for the several examples considered before and discuss how the truncated models compare to estimates of the moments obtained from Monte Carlo simulations. All Monte Carlo simulations were carried out using the algorithm described in [1]. Estimates of the moments were obtained by averaging a large number of Monte Carlo simulations. In most plots, we used a sufficiently large number of simulations so that the 99% confidence intervals for the mean cannot be distinguished from the point estimates at the resolution used for the plots. It is worth to emphasize that the results obtained through Monte Carlo simulations required computational efforts orders of magnitude higher than those obtained using the truncated systems.

Example 1 (Random walk): For this system we consider test functions of the form $\psi^{(m)}(x) = x^m$, $\forall m \in \mathbb{N}$, for which

$$\begin{aligned}(L\psi^{(m)})(x) &= -m\, a\, x^m + \frac{(x + b\sqrt{\epsilon})^m + (x - b\sqrt{\epsilon})^m - 2x^m}{2\epsilon} \\ &= -m\, a\, x^m + \sum_{\substack{i=2\\ i \text{ even}}}^{m} \binom{m}{i} b^i \epsilon^{\frac{i-2}{2}} \psi^{(m-i)}(x),\end{aligned}$$

[3]We are considering polynomials with integer (both positive and negative) powers.

yielding

$$\begin{bmatrix}\dot\mu^{(1)}\\ \dot\mu^{(2)}\\ \dot\mu^{(3)}\\ \dot\mu^{(4)}\\ \dot\mu^{(5)}\\ \vdots\end{bmatrix} = \begin{bmatrix}0\\ b\\ 0\\ b^2\epsilon\\ 0\\ \vdots\end{bmatrix} + \begin{bmatrix}-a & 0 & \cdots & & & & \\ 0 & -2a & 0 & \cdots & & & \\ 3b & 0 & -3a & 0 & \cdots & & \\ 0 & 6b & 0 & -4a & 0 & \cdots & \\ 5b^2\epsilon & 0 & 10b & 0 & -5a & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots\end{bmatrix}\begin{bmatrix}\mu^{(1)}\\ \mu^{(2)}\\ \mu^{(3)}\\ \mu^{(4)}\\ \mu^{(5)}\\ \vdots\end{bmatrix}.$$

Since $\mu^{(0)}(t) = 1$, $\forall t \geq 0$ we excluded $\mu^{(0)}$ from the vector μ, resulting in an affine system (as opposed to linear).

Because the infinite-dimensional moment dynamics are lower-triangular, this system can be truncated without introducing any error. For example, a third order truncation yields

$$\begin{bmatrix}\dot\mu^{(1)}\\ \dot\mu^{(2)}\\ \dot\mu^{(3)}\end{bmatrix} = \begin{bmatrix}-a & 0 & 0\\ 0 & -2a & 0\\ 3b & 0 & -3a\end{bmatrix}\begin{bmatrix}\mu^{(1)}\\ \mu^{(2)}\\ \mu^{(3)}\end{bmatrix} + \begin{bmatrix}0\\ b\\ 0\end{bmatrix}$$

which can be solved explicitly:

$$\mu^{(1)}(t) = e^{-at}\mu^{(1)}(0), \quad \mu^{(2)}(t) = \frac{b}{2a}(1 - e^{-2at}) + e^{-2at}\mu^{(2)}(0),$$
$$\mu^{(3)}(t) = \frac{3b}{2a}(e^{-at} - e^{-3at})\mu^{(1)}(0) + e^{-3at}\mu^{(3)}(0),$$

for $a \neq 0$, and

$$\mu^{(1)}(t) = \mu^{(1)}(0), \qquad \mu^{(2)}(t) = bt + \mu^{(2)}(0), \qquad \mu^{(3)}(t) = 3bt\mu^{(1)}(0) + \mu^{(3)}(0),$$

for $a = 0$. It is interesting to note that these equations do not depend on ϵ because this parameter only appears in the equation for $\dot\mu^{(4)}$. This means that up to the third order moments this pSHS is indistinguishable from the SDE $\dot{\mathbf{x}} = a\mathbf{x} + b\dot{\mathbf{n}}$. $\square$

Example 2 (TCP long-lived): For this system it is particularly meaningful to consider moments of the packet sending rate $\mathbf{r} := \frac{\mathbf{w}}{RTT}$. We therefore choose the test functions to be of the form $\psi^{(m)}(w) = \frac{w^m}{RTT^m}$, $\forall m \in \mathbb{N}$, for which

$$(L\psi^{(m)})(w) = \frac{m\,w^{m-1}}{RTT^{m+1}} - \frac{2^m - 1}{2^m}\frac{p\,w^{m+1}}{RTT^{m+1}} = \frac{m\,\psi^{(m-1)}(w)}{RTT^2} - p\frac{2^m - 1}{2^m}\psi^{(m+1)}(w),$$

yielding

$$\begin{bmatrix}\dot\mu^{(1)}\\ \dot\mu^{(2)}\\ \dot\mu^{(3)}\\ \dot\mu^{(4)}\\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{1}{RTT^2}\\ 0\\ 0\\ 0\\ \vdots\end{bmatrix} + \begin{bmatrix}0 & -\frac{p}{2} & 0 & \cdots & & & \\ \frac{2}{RTT^2} & 0 & -\frac{3p}{4} & 0 & \cdots & & \\ 0 & \frac{3}{RTT^2} & 0 & -\frac{7p}{8} & 0 & \cdots & \\ 0 & 0 & \frac{4}{RTT^2} & 0 & -\frac{15p}{16} & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots\end{bmatrix}\begin{bmatrix}\mu^{(1)}\\ \mu^{(2)}\\ \mu^{(3)}\\ \mu^{(4)}\\ \vdots\end{bmatrix}.$$

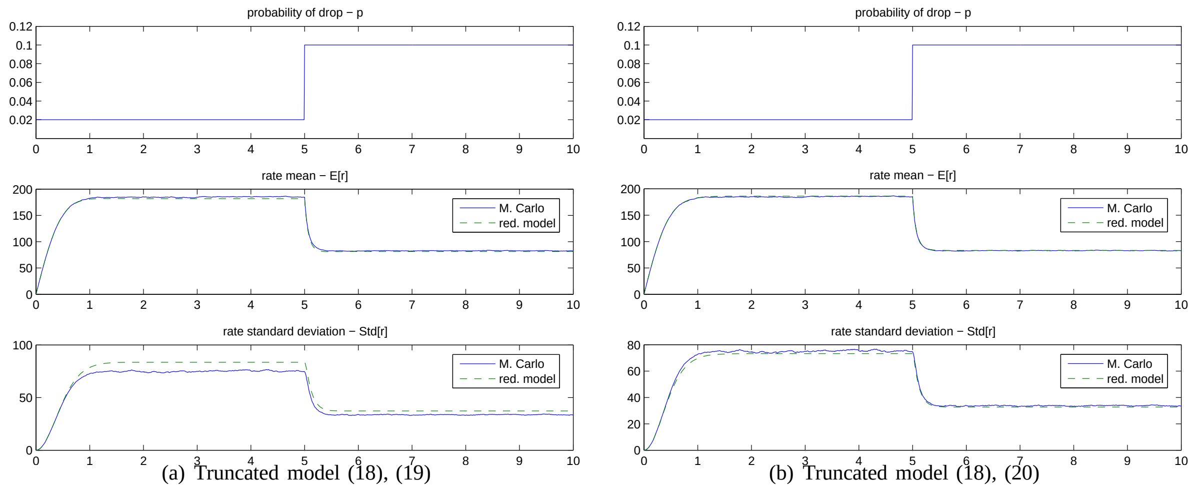


Fig. 8. Comparison between Monte Carlo simulations and two truncated models for Example 2, with $RTT = 50$ms and a step in the drop-rate p from 2% to 10% at time $t = 5$sec.

Since $\mu^{(0)}(t) = 1$, $\forall t \geq 0$ we excluded $\mu^{(0)}$ from the vector μ, resulting in an affine system (as opposed to linear).

We consider a truncation whose state contains the first and second moments of the sending rate. In this case, (9) can be written as follows:

$$\begin{bmatrix} \dot{\mu}^{(1)} \\ \dot{\mu}^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{p}{2} \\ \frac{2}{RTT^2} & 0 \end{bmatrix} \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix} + \begin{bmatrix} \frac{1}{RTT^2} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{3p}{4} \end{bmatrix} \bar{\mu}, \tag{18}$$

where $\bar{\mu} := \mu^{(3)}$ evolves according to $\dot{\mu}^{(3)} = \frac{3\mu^{(2)}}{RTT^2} - \frac{7p\mu^{(4)}}{8}$. In this case, (17) has a unique solution φ, resulting in a truncated system given by (18) and

$$\bar{\mu} = \varphi(\mu^{(1)}, \mu^{(2)}) := \frac{(\mu^{(2)})^{\frac{5}{2}}}{(\mu^{(1)})^2}. \tag{19}$$

Figure 8(a) shows a comparison between Monte Carlo simulations and this truncated model. The dynamics of the first moment of the sending rate is very accurately predicted by the truncated model and there is roughly a 10% error in the steady-state value of the standard deviation. In the simulations, a step in the drop probability was introduced at time $t = 5$sec to show that the truncated model also captures well transient behavior. Figure 8(b) refers to a somewhat more accurate truncated model proposed in [1], which corresponds to

$$\bar{\mu} = \varphi(\mu^{(1)}, \mu^{(2)}) := \left(\frac{\mu^{(2)}}{\mu^{(1)}}\right)^3. \tag{20}$$

Its construction relied on the experimental observation that the steady-state distribution of the sending rate appears to be approximately Log-Normal. It turns out that this model could also have been obtained using the procedure described in Sec. IV by requiring (13) to hold only for $N = 0$ but for a set Ω_μ consisting of all Log-Normal distributions, which is richer than the set $\mathcal{F}_{\text{det}}$ of deterministic distributions used to construct (19). $\square$

Example 3 (TCP on-off): For this system we consider moments of the sending rate $\mathbf{r} := \frac{\mathbf{w}}{RTT}$ on the ss and ca modes, and therefore we use the following test functions

$$\psi^{(0)}_{\mathrm{off}}(q,w) = \begin{cases} 1 & q = \mathrm{off} \\ 0 & q \in \{\mathrm{ca}, \mathrm{ss}\} \end{cases} \qquad \psi^{(m)}_{\mathrm{ss}}(q,w) = \begin{cases} \frac{w^m}{RTT^m} & q = \mathrm{ss} \\ 0 & q \in \{\mathrm{ca}, \mathrm{off}\} \end{cases} \qquad \psi^{(m)}_{\mathrm{ca}}(q,w) = \begin{cases} \frac{w^m}{RTT^m} & q = \mathrm{ca} \\ 0 & q \in \{\mathrm{ss}, \mathrm{off}\}, \end{cases}$$

for which

$$\begin{aligned}
(L\psi^{(0)}_{\mathrm{off}})(q,w) &= -\tau^{-1}_{\mathrm{off}}\psi^{(0)}_{\mathrm{off}}(q,w) + \frac{1}{k}\psi^{(1)}_{\mathrm{ss}}(q,w) + \frac{1}{k}\psi^{(1)}_{\mathrm{ca}}(q,w) \\
(L\psi^{(m)}_{\mathrm{ss}})(q,w) &= \frac{\tau^{-1}_{\mathrm{off}}\, w_0^m}{RTT^m}\psi^{(0)}_{\mathrm{off}}(q,w) + \frac{m\log 2}{RTT}\psi^{(m)}_{\mathrm{ss}}(q,w) - \Big(p + \frac{1}{k}\Big)\psi^{(m+1)}_{\mathrm{ss}}(q,w) \\
(L\psi^{(m)}_{\mathrm{ca}})(q,w) &= \frac{p}{2^m}\,\psi^{(m+1)}_{\mathrm{ss}}(q,w) + \frac{m}{RTT^2}\psi^{(m-1)}_{\mathrm{ca}}(q,w) - \Big(\frac{1}{k} + p - \frac{p}{2^m}\Big)\psi^{(m+1)}_{\mathrm{ca}}(q,w),
\end{aligned}$$

yielding

$$\begin{bmatrix} \dot\mu^{(0)}_{\mathrm{off}} \\ \dot\mu^{(0)}_{\mathrm{ss}} \\ \dot\mu^{(0)}_{\mathrm{ca}} \\ \dot\mu^{(1)}_{\mathrm{ss}} \\ \dot\mu^{(1)}_{\mathrm{ca}} \\ \dot\mu^{(2)}_{\mathrm{ss}} \\ \dot\mu^{(2)}_{\mathrm{ca}} \\ \dot\mu^{(3)}_{\mathrm{ss}} \\ \dot\mu^{(3)}_{\mathrm{ca}} \\ \vdots \end{bmatrix} = \begin{bmatrix}
-\tau^{-1}_{\mathrm{off}} & 0 & 0 & \frac{1}{k} & \frac{1}{k} & 0 & 0 & 0 & 0 & \cdots & & & \\
\tau^{-1}_{\mathrm{off}} & 0 & 0 & -\frac{1}{k}-p & 0 & 0 & 0 & 0 & 0 & \cdots & & & \\
0 & 0 & 0 & p & -\frac{1}{k} & 0 & 0 & 0 & 0 & \cdots & & & \\
\frac{\tau^{-1}_{\mathrm{off}} w_0}{RTT} & 0 & 0 & \frac{\log 2}{RTT} & 0 & -\frac{1}{k}-p & 0 & 0 & 0 & \cdots & & & \\
0 & 0 & \frac{1}{RTT^2} & 0 & 0 & \frac{p}{2} & -\frac{1}{k}-\frac{p}{2} & 0 & 0 & \cdots & & & \\
\frac{\tau^{-1}_{\mathrm{off}} w_0^2}{RTT^2} & 0 & 0 & 0 & 0 & \frac{\log 4}{RTT} & 0 & -\frac{1}{k}-p & 0 & \cdots & & & \\
0 & 0 & 0 & 0 & \frac{2}{RTT^2} & 0 & 0 & \frac{p}{4} & -\frac{1}{k}-\frac{3p}{4} & 0 & \cdots & & \\
\frac{\tau^{-1}_{\mathrm{off}} w_0^3}{RTT^3} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\log 8}{RTT} & 0 & -\frac{1}{k}-p & 0 & \cdots & \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{RTT^2} & 0 & 0 & \frac{p}{8} & -\frac{1}{k}-\frac{7p}{8} & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots &
\end{bmatrix} \begin{bmatrix} \mu^{(0)}_{\mathrm{off}} \\ \mu^{(0)}_{\mathrm{ss}} \\ \mu^{(0)}_{\mathrm{ca}} \\ \mu^{(1)}_{\mathrm{ss}} \\ \mu^{(1)}_{\mathrm{ca}} \\ \mu^{(2)}_{\mathrm{ss}} \\ \mu^{(2)}_{\mathrm{ca}} \\ \mu^{(3)}_{\mathrm{ss}} \\ \mu^{(3)}_{\mathrm{ca}} \\ \mu^{(4)}_{\mathrm{ss}} \\ \mu^{(4)}_{\mathrm{ca}} \\ \vdots \end{bmatrix}.$$

Since $\mu^{(0)}_{\mathrm{off}}$, $\mu^{(0)}_{\mathrm{ss}}$, and $\mu^{(0)}_{\mathrm{ca}}$ are the probabilities that $\mathbf{q}$ is equal to off, ss, and ca, respectively, these quantities must add up to one. We could therefore exclude one of them from μ.

We consider a truncation whose state contains the zeroth, first, and second moments of the sending rate. In this case, (9) can be written as follows:

$$\begin{bmatrix} \dot\mu^{(0)}_{\mathrm{off}} \\ \dot\mu^{(0)}_{\mathrm{ss}} \\ \dot\mu^{(0)}_{\mathrm{ca}} \\ \dot\mu^{(1)}_{\mathrm{ss}} \\ \dot\mu^{(1)}_{\mathrm{ca}} \\ \dot\mu^{(2)}_{\mathrm{ss}} \\ \dot\mu^{(2)}_{\mathrm{ca}} \end{bmatrix} = \begin{bmatrix}
-\tau^{-1}_{\mathrm{off}} & 0 & 0 & \frac{1}{k} & \frac{1}{k} & 0 & 0 \\
\tau^{-1}_{\mathrm{off}} & 0 & 0 & -\frac{1}{k}-p & 0 & 0 & 0 \\
0 & 0 & 0 & p & -\frac{1}{k} & 0 & 0 \\
\frac{\tau^{-1}_{\mathrm{off}} w_0}{RTT} & 0 & 0 & \frac{\log 2}{RTT} & 0 & -\frac{1}{k}-p & 0 \\
0 & 0 & \frac{1}{RTT^2} & 0 & 0 & \frac{p}{2} & -\frac{1}{k}-\frac{p}{2} \\
\frac{\tau^{-1}_{\mathrm{off}} w_0^2}{RTT^2} & 0 & 0 & 0 & 0 & \frac{\log 4}{RTT} & 0 \\
0 & 0 & 0 & 0 & \frac{2}{RTT^2} & 0 & 0
\end{bmatrix} \begin{bmatrix} \mu^{(0)}_{\mathrm{off}} \\ \mu^{(0)}_{\mathrm{ss}} \\ \mu^{(0)}_{\mathrm{ca}} \\ \mu^{(1)}_{\mathrm{ss}} \\ \mu^{(1)}_{\mathrm{ca}} \\ \mu^{(2)}_{\mathrm{ss}} \\ \mu^{(2)}_{\mathrm{ca}} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -\frac{1}{k}-p & 0 \\ \frac{p}{4} & -\frac{1}{k}-\frac{3p}{4} \end{bmatrix} \bar\mu, \quad (21)$$

where $\bar\mu := \begin{bmatrix} \mu^{(3)}_{\mathrm{ss}} & \mu^{(3)}_{\mathrm{ca}} \end{bmatrix}'$ evolves according to

$$\dot\mu^{(3)}_{\mathrm{ss}} = \frac{\tau^{-1}_{\mathrm{off}} w_0^3}{RTT^3}\mu^{(0)}_{\mathrm{off}} + \frac{\log 8}{RTT}\mu^{(3)}_{\mathrm{ss}} - \Big(\frac{1}{k} + p\Big)\mu^{(4)}_{\mathrm{ss}}, \quad (22a)$$

$$\dot\mu^{(3)}_{\mathrm{ca}} = \frac{3}{RTT^2}\mu^{(2)}_{\mathrm{ca}} + \frac{p}{8}\mu^{(4)}_{\mathrm{ss}} - \Big(\frac{1}{k} + \frac{7p}{8}\Big)\mu^{(4)}_{\mathrm{ca}}. \quad (22b)$$

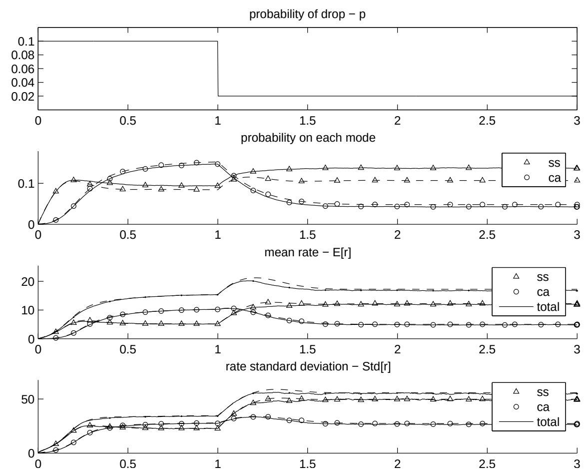


Fig. 9. (left) Comparison between Monte Carlo simulations (solid lines) and the truncated model (21), (24) (dashed lines) for Example 3, with $RTT = 50$ms, $\tau_{\text{off}} = 1$sec, $k = 20.39$ packets (corresponding to 30.58KB files broken into 1500bytes packets, which is consistent with the file-size distribution of the UNIX file system reported in [16]), and a step in the drop-rate p from 10% to 2% at $t = 1$sec.

However, (22) does not satisfy condition 5 in Lemma 1 because the different discrete modes do not appear decoupled: $\dot{\mu}_{\text{ss}}^{(3)}$ depends on $\mu_{\text{off}}^{(0)}$, and $\dot{\mu}_{\text{ca}}^{(3)}$ depends on $\mu_{\text{ss}}^{(4)}$. For the purpose of determining φ, we ignore the cross coupling terms and approximate (22) by

$$\dot{\mu}_{\text{ss}}^{(3)} \approx \frac{\log 8}{RTT}\mu_{\text{ss}}^{(3)} - \Big(\frac{1}{k} + p\Big)\mu_{\text{ss}}^{(4)}, \qquad \dot{\mu}_{\text{ca}}^{(3)} \approx \frac{3}{RTT^2}\mu_{\text{ca}}^{(2)} - \Big(\frac{1}{k} + \frac{7p}{8}\Big)\mu_{\text{ca}}^{(4)}. \tag{23}$$

The validity of these approximations generally depends on the network parameters. When (23) is used, it is straightforward to verify that (17) has a unique solution φ, resulting in a truncated system given by (21) and

$$\bar{\mu} = \varphi(\mu) = \begin{bmatrix} \frac{\mu_{\text{ss}}^{(0)}(\mu_{\text{ss}}^{(2)})^3}{(\mu_{\text{ss}}^{(1)})^3} & \frac{(\mu_{\text{ca}}^{(0)})^{\frac{1}{2}}(\mu_{\text{ca}}^{(2)})^{\frac{5}{2}}}{(\mu_{\text{ca}}^{(1)})^2} \end{bmatrix}'. \tag{24}$$

Since $\mu_{\text{off}}^{(0)}$, $\mu_{\text{ss}}^{(0)}$, and $\mu_{\text{ca}}^{(0)}$ correspond to the probabilities that $\mathbf{q}$ is equal to off, ss, and ca, respectively, these quantities must add up to one and we can exclude one of them from the state.

Figure 9 shows a comparison between Monte Carlo simulations and the truncated model (21), (24). The dynamics of the first and second order moments are accurately predicted by the truncated model. There is some steady-state error in the probabilities of each mode, especially after the drop-rate p dropped to 2%. This corresponds to a situation in which the sending rate in the ss mode exceeds that of the ca mode and therefore the $\mu_{\text{ss}}^{(4)}$ term dropped from (23) in the $\dot{\mu}_{\text{ca}}^{(3)}$ equation may actually have a significant effect. Somewhat surprisingly, even in this case the first and second order moments are very well approximated. In preparing this paper, several simulation were executed for different network parameters and initial conditions. Figure 9 shows typical best-case (before $t = 1$) and worst-case (after $t = 1$) results.

The average file size used to produce Fig. 9 is consistent with the one observed in the UNIX file system reported in [16]. It should be emphasized that the overall distribution of the UNIX file system and more importantly of files transmitted over the Internet is not well approximated by an exponential distribution. More reasonable distributions are Pareto or mixtures of exponentials. Both could be modeled by pSHSs, but for simplicity here we focused our attention on the simplest case of a single exponential distribution. □

Example 4 (Networked control system): For this system we consider test functions of the form $\psi^{(m)}(e) = e^m$, $\forall m \in \mathbb{N}$, for which

$$\begin{aligned}(L\psi^{(m)})(e) &= a\,m\,e^m + \frac{m(m-1)b^2}{2}e^{m-2} - e^{m+2\rho}\\ &= a\,m\,\psi^{(m)}(e) + \frac{m(m-1)b^2}{2}\psi^{(m-2)}(e) - \psi^{(m+2\rho)}(e),\end{aligned}$$

yielding

$$\begin{bmatrix}\dot\mu^{(1)}\\ \dot\mu^{(2)}\\ \dot\mu^{(3)}\\ \dot\mu^{(4)}\\ \dot\mu^{(5)}\\ \vdots\end{bmatrix} = \begin{bmatrix} a\mu^{(1)} - \mu^{(1+2\rho)}\\ b^2 + 2a\mu^{(2)} - \mu^{(2+2\rho)}\\ 3b^2\mu^{(1)} + 3a\mu^{(3)} - \mu^{(3+2\rho)}\\ 6b^2\mu^{(2)} + 4a\mu^{(4)} - \mu^{(4+2\rho)}\\ 10b^2\mu^{(3)} + 5a\mu^{(5)} - \mu^{(5+2\rho)}\\ \vdots\end{bmatrix}.$$

In particular, for $\rho = 0$ we get

$$\begin{bmatrix}\dot\mu^{(1)}\\ \dot\mu^{(2)}\\ \dot\mu^{(3)}\\ \dot\mu^{(4)}\\ \dot\mu^{(5)}\\ \vdots\end{bmatrix} = \begin{bmatrix}0\\ -b^2\\ 0\\ 0\\ 0\\ \vdots\end{bmatrix} + \begin{bmatrix} a-1 & 0 & \cdots & & & & \\ 0 & 2a-1 & 0 & \cdots & & & \\ -3b^2 & 0 & 3a-1 & 0 & \cdots & & \\ 0 & -6b^2 & 0 & 4a-1 & 0 & \cdots & \\ 0 & 0 & -10b^2 & 0 & 5a-1 & 0 & \cdots\\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots\end{bmatrix}\begin{bmatrix}\mu^{(1)}\\ \mu^{(2)}\\ \mu^{(3)}\\ \mu^{(4)}\\ \mu^{(5)}\\ \vdots\end{bmatrix}$$

and for $\rho = 1$ we get

$$\begin{bmatrix}\dot\mu^{(1)}\\ \dot\mu^{(2)}\\ \dot\mu^{(3)}\\ \dot\mu^{(4)}\\ \dot\mu^{(5)}\\ \vdots\end{bmatrix} = \begin{bmatrix}0\\ -b^2\\ 0\\ 0\\ 0\\ \vdots\end{bmatrix} + \begin{bmatrix} a & 0 & -1 & 0 & \cdots & & & & \\ 0 & 2a & 0 & -1 & 0 & \cdots & & & \\ -3b^2 & 0 & 3a & 0 & -1 & 0 & \cdots & & \\ 0 & -6b^2 & 0 & 4a & 0 & -1 & 0 & \cdots & \\ 0 & 0 & -10b^2 & 0 & 5a & 0 & -1 & 0 & \cdots\\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots\end{bmatrix}\begin{bmatrix}\mu^{(1)}\\ \mu^{(2)}\\ \mu^{(3)}\\ \mu^{(4)}\\ \mu^{(5)}\\ \vdots\end{bmatrix}.$$

Since $\mu^{(0)}(t) = 1$, $\forall t \geq 0$ we excluded $\mu^{(0)}$ from the vector μ, resulting in an affine system (as opposed to linear).

For $\rho = 0$, the infinite-dimensional dynamics have a lower-triangular structure and therefore an exact truncation is possible. However, this case is less interesting because it corresponds to a reset-rate that does not depend on the continuous state and is therefore farther from the optimal [9, 10]. We consider here $\rho = 1$. In this case, the odd and even moments are decoupled and can be studied independently. It is straightforward to check that if the

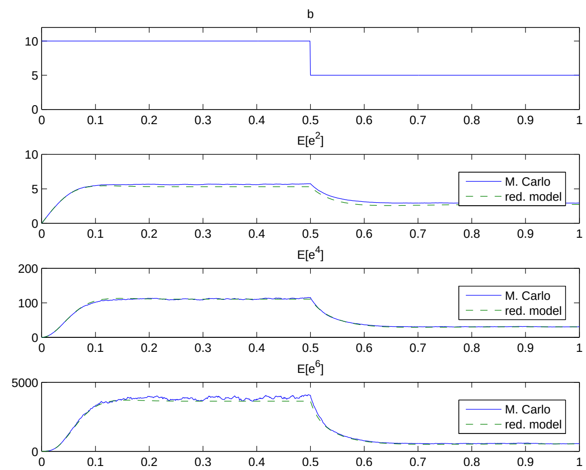


Fig. 10. (right) Comparison between Monte Carlo simulations (solid lines) and the truncated model (25), (26) (dashed lines) for Example 4, with $a = 1$, $q = 1$, and a step in the parameter b from 10 to 2 at time $t = 0.5$sec.

initial distribution of $\mathbf{e}$ is symmetric around zero, it will remain so for all times and therefore all odd moments are constant and equal to zero. Regarding the even moments, the smallest truncation for which condition 4 in Lemma 1 holds is a third order one, for which (9) can be written as follows:

$$\begin{bmatrix} \dot{\mu}^{(2)} \\ \dot{\mu}^{(4)} \\ \dot{\mu}^{(6)} \end{bmatrix} = \begin{bmatrix} 2a & -1 & 0 \\ 6b^2 & 4a & -1 \\ 0 & 15b^2 & 6a \end{bmatrix} \begin{bmatrix} \mu^{(2)} \\ \mu^{(4)} \\ \mu^{(6)} \end{bmatrix} + \begin{bmatrix} b^2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \bar{\mu}, \tag{25}$$

where $\bar{\mu} := \mu^{(8)}$ evolves according to $\dot{\mu}^{(8)} = -28b^2\mu^{(6)} + 8a\mu^{(8)} - \mu^{(10)}$. It is straightforward to verify that (17) has a unique solution φ, resulting in a truncated system given by (25) and

$$\bar{\mu} = \varphi(\mu^{(2)}, \mu^{(4)}, \mu^{(6)}) = \mu^{(2)} \left(\frac{\mu^{(6)}}{\mu^{(4)}} \right)^3. \tag{26}$$

Figure 10 shows a comparison between Monte Carlo simulations and the truncated model (25), (26). The dynamics of the all the moments are accurately predicted by the truncated model. The nonlinearity of the underlying model is apparent by the fact that halving b at time $t = 0.5$sec, which corresponds to dividing the variance of the noise by 4, only results in approximately dividing the variance of the estimation error by 2. □

Example 5 (Decaying reaction): For this system we consider test functions of the form $\psi^{(m)}(x) = x^m$, $\forall m \in \mathbb{N}$, for which

$$(L\psi^{(m)})(x) = \frac{c}{2}x(x-1)\big((x-2)^m - x^m\big) = -\frac{c}{2}\bigg(-\psi^{(m+1)} + \sum_{k=0}^{m} \binom{m}{k} \frac{m+k+1}{m-k+1} (-2)^{m-k} \psi^{(k+1)} \bigg), \tag{27}$$

yielding

$$\begin{bmatrix} \dot{\mu}^{(1)} \\ \dot{\mu}^{(2)} \\ \dot{\mu}^{(3)} \\ \dot{\mu}^{(4)} \\ \dot{\mu}^{(5)} \\ \vdots \end{bmatrix} = c \begin{bmatrix} 1 & -1 & 0 & \cdots & & & & \\ -2 & 4 & -2 & 0 & \cdots & & & \\ 4 & -10 & 9 & -3 & 0 & \cdots & & \\ -8 & 24 & -28 & 16 & -4 & 0 & \cdots & \\ 16 & -56 & 80 & -60 & 25 & -5 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \\ \mu^{(3)} \\ \mu^{(4)} \\ \mu^{(5)} \\ \vdots \end{bmatrix}.$$

For this example it is not possible to find a function φ for which (17) holds exactly for any finite truncation. However, finding φ becomes possible if we ignore some of the lowest-order powers of x in the summation in (27) that defines the generator of the system. This is reasonable when the number of particles is significantly larger than one. In the extreme case, for which we only leave the highest order term, a first-order truncation is possible and (9) can be written as follows:

$$\dot{\mu}^{(1)} = -c\bar{\mu},$$

where $\bar{\mu} := \mu^{(2)}$ evolves according to $\dot{\mu}^{(2)} = -2c\mu^{(3)}$. In this case, (17) has a unique solution φ, resulting in the following truncated system

$$\dot{\nu}^{(1)} = -c\,(\nu^{(1)})^2,$$

which would also be obtained using the deterministic formulation of chemical kinetics. However, by leaving more terms one can obtain approximations that should be more accurate, especially for smaller number of particles. For example, if one would keep the two highest powers of x in the summation in (27), a second-order truncation is possible and (9) can be written as follows:

$$\begin{bmatrix} \dot{\mu}^{(1)} \\ \dot{\mu}^{(2)} \end{bmatrix} = \begin{bmatrix} c & -c \\ 0 & 4c \end{bmatrix} \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix} + \begin{bmatrix} 0 \\ -2c \end{bmatrix} \bar{\mu},$$

where $\bar{\mu} := \mu^{(3)}$ evolves according to $\dot{\mu}^{(3)} = 9c\mu^{(3)} - 3c\mu^{(4)}$. In this case, (17) has a unique solution φ, resulting in the following truncated system

$$\begin{bmatrix} \dot{\nu}^{(1)} \\ \dot{\nu}^{(2)} \end{bmatrix} = \begin{bmatrix} c & -c \\ 0 & 4c \end{bmatrix} \begin{bmatrix} \nu^{(1)} \\ \nu^{(2)} \end{bmatrix} + \begin{bmatrix} 0 \\ -2c \end{bmatrix} \left(\frac{\nu^{(2)}}{\nu^{(1)}}\right)^3.$$

The relationship between the first three moments expressed by this φ is exact for Log-Normal distributions. It turns out that if we had kept a larger number of higher-order powers of x in the summation in (27), we would have obtained higher-order truncation for which φ that would also be exactly compatible with Log-Normal distributions. This observation has been further explored in [17]. For conciseness, we omit the comparison between Monte Carlo simulations and the truncated model for this very simple reaction. □

Example 6 (Decaying-dimerizing reaction set): For this system we consider test functions of the form

$$\psi^{(m_1,m_2,m_2)}(x) = x_1^{m_1} x_2^{m_2} x_3^{m_3}, \qquad \forall m_1, m_2, m_3 \in \mathbb{N},$$

for which

$$
\begin{aligned}
(L\psi^{(m_1,m_2,m_2)})(x) &= c_1 x_1\big((x_1-1)^{m_1} - x_1^{m_1}\big)x_2^{m_2}x_3^{m_3} \\
&\quad + \frac{c_2}{2}x_1(x_1-1)\big((x_1-2)^{m_1}(x_2+1)^{m_2} - x_1^{m_1}x_2^{m_2}\big)x_3^{m_3} \\
&\quad + c_3 x_2\big((x_1+2)^{m_1}(x_2-1)^{m_2} - x_1^{m_1}x_2^{m_2}\big)x_3^{m_3} + c_4 x_2\big((x_2-1)^{m_2}(x_3+1)^{m_3} - x_2^{m_2}x_3^{m_3}\big)x_1^{m_1} \\
&= c_1 \sum_{i=0}^{m_1-1} \binom{m_1}{i}(-1)^{m_1-i}\psi^{(i+1,m_2,m_3)}(x) \\
&\quad + \frac{c_2}{2} \sum_{\substack{i,j=0 \\ (i,j)\neq(m_1,m_2)}}^{m_1,m_2} \binom{m_1}{i}\binom{m_2}{j}(-2)^{m_1-i}\big(\psi^{(i+2,j,m_3)}(x) - \psi^{(i+1,j,m_3)}(x)\big) \\
&\quad + c_3 \sum_{\substack{i,j=0 \\ (i,j)\neq(m_1,m_2)}}^{m_1,m_2} \binom{m_1}{i}\binom{m_2}{j} 2^{m_1-i}(-1)^{m_2-j}\psi^{(i,j+1,m_3)}(x) \\
&\quad + c_4 \sum_{\substack{i,j=0 \\ (i,j)\neq(m_2,m_3)}}^{m_2,m_3} \binom{m_2}{i}\binom{m_3}{j}(-1)^{m_2-i}\psi^{(m_1,i+1,j)}(x),
\end{aligned}
\tag{28}
$$

where the summations result from the power expansions of the terms $(x_i - c)^{m_i}$. Therefore

$$
\begin{bmatrix} \dot\mu^{(1,0,0)} \\ \dot\mu^{(0,1,0)} \\ \dot\mu^{(0,0,1)} \\ \dot\mu^{(2,0,0)} \\ \dot\mu^{(0,2,0)} \\ \dot\mu^{(1,1,0)} \\ \dot\mu^{(3,0,0)} \\ \dot\mu^{(2,1,0)} \\ \vdots \end{bmatrix}
=
\begin{bmatrix}
-c_1+c_2 & 2c_3 & 0 & -c_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
-\frac{c_2}{2} & -c_3-c_4 & 0 & \frac{c_2}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & c_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_1-2c_2 & 4c_3 & 0 & -2c_1+4c_2 & 0 & 4c_3 & -2c_2 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
-\frac{c_2}{2} & c_3+c_4 & 0 & \frac{c_2}{2} & -2c_3-2c_4 & -c_2 & 0 & c_2 & 0 & 0 & 0 & 0 & 0 & \cdots \\
c_2 & -2c_3 & 0 & -\frac{3c_2}{2} & 2c_3 & -c_1+c_2-c_3-c_4 & \frac{c_2}{2} & -c_2 & 0 & 0 & 0 & 0 & 0 & \cdots \\
-c_1+4c_2 & 8c_3 & 0 & 3c_1-10c_2 & 0 & 12c_3 & -3c_1+9c_2 & 6c_3 & 0 & -3c_2 & 0 & 0 & 0 & \cdots \\
-2c_2 & -4c_3 & 0 & 4c_2 & 4c_3 & c_1-2c_2-4c_3 & -\frac{5c_2}{2} & 0 & -2c_1+4c_2-c_3-c_4 & 4c_3 & \frac{c_2}{2} & -2c_3 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\begin{bmatrix} \mu^{(1,0,0)} \\ \mu^{(0,1,0)} \\ \mu^{(0,0,1)} \\ \mu^{(2,0,0)} \\ \mu^{(0,2,0)} \\ \mu^{(1,1,0)} \\ \mu^{(3,0,0)} \\ \mu^{(2,1,0)} \\ \mu^{(1,2,0)} \\ \mu^{(4,0,0)} \\ \mu^{(3,1,0)} \\ \vdots \end{bmatrix}.
$$

For this example we consider a truncation whose state contains all the first and second order moments for the number of particles of the first and second species. To keep the formulas small, we omit from the truncation the second moments of the third species, which does not appear as a reactant in any reaction and therefore its higher order statistics do not affect the first two. In this case, (9) can be written as follows:

$$
\begin{bmatrix} \dot\mu^{(1,0,0)} \\ \dot\mu^{(0,1,0)} \\ \dot\mu^{(0,0,1)} \\ \dot\mu^{(2,0,0)} \\ \dot\mu^{(0,2,0)} \\ \dot\mu^{(1,1,0)} \end{bmatrix}
=
\begin{bmatrix}
-c_1+c_2 & 2c_3 & 0 & -c_2 & 0 & 0 \\
-\frac{c_2}{2} & -c_3-c_4 & 0 & \frac{c_2}{2} & 0 & 0 \\
0 & c_4 & 0 & 0 & 0 & 0 \\
c_1-2c_2 & 4c_3 & 0 & 4c_2-2c_1 & 0 & 4c_3 \\
-\frac{c_2}{2} & c_3+c_4 & 0 & \frac{c_2}{2} & -2c_3-2c_4 & -c_2 \\
c_2 & -2c_3 & 0 & -\frac{3c_2}{2} & 2c_3 & c_2-c_1-c_3-c_4
\end{bmatrix}
\begin{bmatrix} \mu^{(1,0,0)} \\ \mu^{(0,1,0)} \\ \mu^{(0,0,1)} \\ \mu^{(2,0,0)} \\ \mu^{(0,2,0)} \\ \mu^{(1,1,0)} \end{bmatrix}
+
\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -2c_2 & 0 \\ 0 & c_2 \\ \frac{c_2}{2} & -c_2 \end{bmatrix}
\bar\mu,
\tag{29}
$$

where $\bar\mu := \begin{bmatrix} \mu^{(3,0,0)} & \mu^{(2,1,0)} \end{bmatrix}'$ evolves according to

$$
\begin{aligned}
\dot\mu^{(3,0,0)} &= (-c_1+4c_2)\mu^{(1,0,0)} + 8c_3\mu^{(0,1,0)} + (3c_1-10c_2)\mu^{(2,0,0)} + 12c_3\mu^{(1,1,0)} \\
&\quad + (-3c_1+9c_2)\mu^{(3,0,0)} + 6c_3\mu^{(2,1,0)} - 3c_2\mu^{(4,0,0)}
\end{aligned}
\tag{30a}
$$

$$
\begin{aligned}
\dot\mu^{(2,1,0)} &= -2c_2\mu^{(1,0,0)} - 4c_3\mu^{(0,1,0)} + 4c_2\mu^{(2,0,0)} + 4c_3\mu^{(0,2,0)} + (c_1-2c_2-4c_3)\mu^{(1,1,0)} \\
&\quad - \frac{5c_2\mu^{(3,0,0)}}{2} + (4c_2-2c_1-c_3-c_4)\mu^{(2,1,0)} + 4c_3\mu^{(1,2,0)} + \frac{c_2\mu^{(4,0,0)}}{2} - 2c_2\mu^{(3,1,0)}.
\end{aligned}
\tag{30b}
$$

This system does not satisfy condition 4 in Lemma 1 because the $\mu^{(1,0,0)}$, $\mu^{(0,1,0)}$ terms in the right-hand sides of (30) lead to monomials in x_1 and x_2 in $\sum_{i=1}^{\infty} a_{\ell,i}\, x^{(m_i)}$ that do not exist in any of the polynomials

TABLE I

COMPARISON BETWEEN ESTIMATES OBTAINED FROM MONTE CARLO SIMULATIONS AND THE TRUNCATED MODEL FOR EXAMPLE 6. THE MONTE CARLO SIMULATION DATA WAS TAKEN FROM [13].

Source for the estimates	$\mathrm{E}[\mathbf{x}_1(0.2)]$	$\mathrm{E}[\mathbf{x}_2(0.2)]$	$\mathrm{StdDev}[\mathbf{x}_1(0.2)]$	$\mathrm{StdDev}[\mathbf{x}_2(0.2)]$
10,000 MC. simul.	387.3	749.5	18.42	10.49
model (29), (32)	387.2	749.6	18.54	10.60

$\big\{\sum_{i=1}^{\infty} a_{j,i}\, x^{(m_\ell - m_j + m_i)} : 1 \le j \le 6\big\}$. These terms can be easily traced back to the lowest-order terms in power expansions in (28) and disappear if we only keep the three highest powers of x_1 in the expansion of $(x_1-1)^{m_1}$ that corresponds to the first reaction and the two highest powers of x_1 and x_2 in the expansions of $(x_1 \pm 2)^{m_1}$ and $(x_2 \pm 1)^{m_2}$ that correspond to the second and third reactions. In practice, this leads to the following simplified version of (29)

$$\begin{bmatrix} \dot\mu^{(1,0,0)} \\ \dot\mu^{(0,1,0)} \\ \dot\mu^{(0,0,1)} \\ \dot\mu^{(2,0,0)} \\ \dot\mu^{(0,2,0)} \\ \dot\mu^{(1,1,0)} \end{bmatrix} = \begin{bmatrix} -c_1+c_2 & 2c_3 & 0 & -c_2 & 0 & 0 \\ -\frac{c_2}{2} & -c_3-c_4 & 0 & \frac{c_2}{2} & 0 & 0 \\ 0 & c_4 & 0 & 0 & 0 & 0 \\ c_1 & 0 & 0 & 4c_2-2c_1 & 0 & 4c_3 \\ 0 & c_4 & 0 & 0 & -2c_3-2c_4 & -c_2 \\ c_2 & -2c_3 & 0 & -\frac{3c_2}{2} & 2c_3 & c_2-c_1-c_3-c_4 \end{bmatrix} \begin{bmatrix} \mu^{(1,0,0)} \\ \mu^{(0,1,0)} \\ \mu^{(0,0,1)} \\ \mu^{(2,0,0)} \\ \mu^{(0,2,0)} \\ \mu^{(1,1,0)} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -2c_2 & 0 \\ 0 & c_2 \\ \frac{c_2}{2} & -c_2 \end{bmatrix} \bar\mu, \tag{31}$$

where now $\bar\mu := \begin{bmatrix} \mu^{(3,0,0)} & \mu^{(2,1,0)} \end{bmatrix}'$ evolves according to

$$\dot\mu^{(3,0,0)} = 3c_1\mu^{(2,0,0)} + (-3c_1+3c_2)\mu^{(3,0,0)} + 6c_3\mu^{(2,1,0)} - 3c_2\mu^{(4,0,0)}$$

$$\begin{aligned}\dot\mu^{(2,1,0)} &= 2c_2\mu^{(2,0,0)} + (c_1-4c_3)\mu^{(1,1,0)} - \frac{5c_2\mu^{(3,0,0)}}{2} \\ &\quad + (-2c_1+2c_2-c_3-c_4)\mu^{(2,1,0)} + 4c_3\mu^{(1,2,0)} + \frac{c_2\mu^{(4,0,0)}}{2} - 2c_2\mu^{(3,1,0)},\end{aligned}$$

for which condition 4 in Lemma 1 does hold, allowing us to find a unique solution φ to (17), resulting in a truncated system given by (29) and

$$\bar\mu = \varphi(\mu) = \left[\left(\frac{\mu^{(2,0,0)}}{\mu^{(1,0,0)}}\right)^3 \quad \frac{\mu^{(2,0,0)}}{\mu^{(0,1,0)}} \left(\frac{\mu^{(1,1,0)}}{\mu^{(1,0,0)}}\right)^2 \right]'. \tag{32}$$

Ignoring the lowest-order powers of x_1 and x_2 in the power expansions is valid when the populations of these species are high. In practice, the approximation still seems to yield good results even when the populations are fairly small as shown by the Monte Carlo simulations.

Figure 11 shows a comparison between Monte Carlo simulations and the truncated model (29), (32). The coefficients used were taken from [13, Example 1]: $c_1 = 1$, $c_2 = 10$, $c_3 = 1000$, $c_4 = 10^{-1}$. In Fig. 11(a), we used the same initial conditions as in [13, Example 1]:

$$\mathbf{x}_1(0) = 400, \qquad \mathbf{x}_2(0) = 798, \qquad \mathbf{x}_3(0) = 0. \tag{33}$$

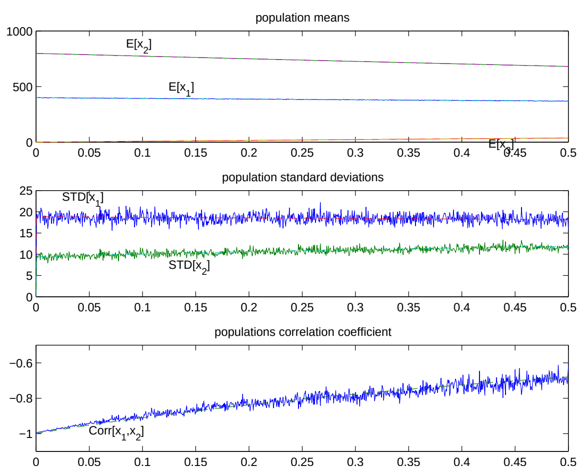
The match is very accurate, as can be confirmed from Table I. The values of the parameters chosen result in a pSHS with two distinct time scales, which makes this pSHS computationally difficult to simulate ("stiff" in the terminology of [13]). The initial conditions (33) start in the "slow manifold" $\mathbf{x}_2 = \frac{5}{1000}\mathbf{x}_1(\mathbf{x}_1 - 1)$ and Fig. 11(a) essentially shows the evolution of the system on this manifold. Figure 11(b) zooms in on the interval $[0, 5 \times 10^{-4}]$ and shows the evolution of the system towards this manifold when it starts away from it at

$$\mathbf{x}_1(0) = 800, \qquad \mathbf{x}_2(0) = 100, \qquad \mathbf{x}_3(0) = 200. \tag{34}$$

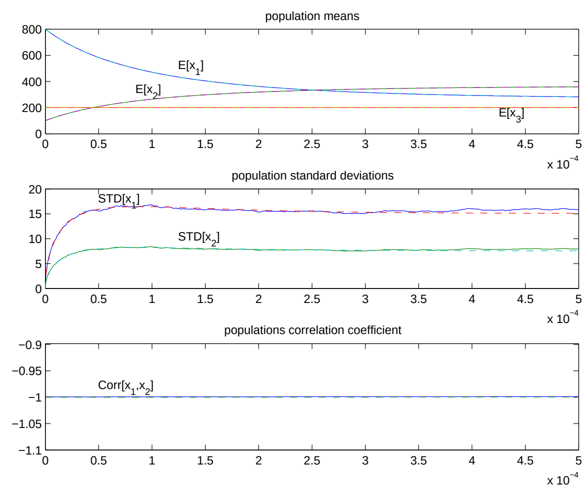
Figures 11(c)–11(d) shows another simulation of the same reactions but for much smaller initial populations:

$$\mathbf{x}_1(0) = 10, \qquad \mathbf{x}_2(0) = 10, \qquad \mathbf{x}_3(0) = 5. \tag{35}$$

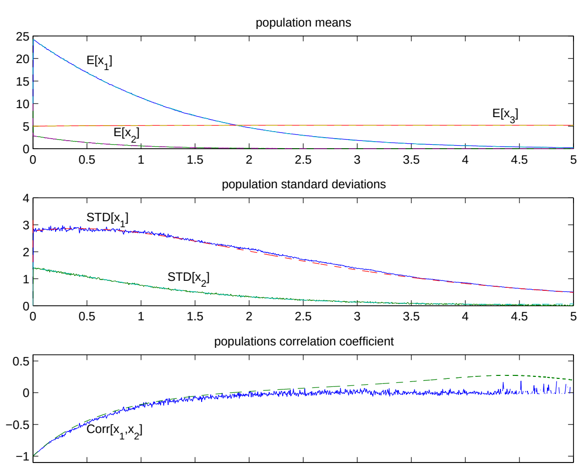
The truncated model still provides an extremely good approximation, with significant error only in the covariance between $\mathbf{x}_1$ and $\mathbf{x}_2$ when the averages and standard deviation of these variables get below one. This happens in spite of having used (31) instead of (30) to compute φ. □



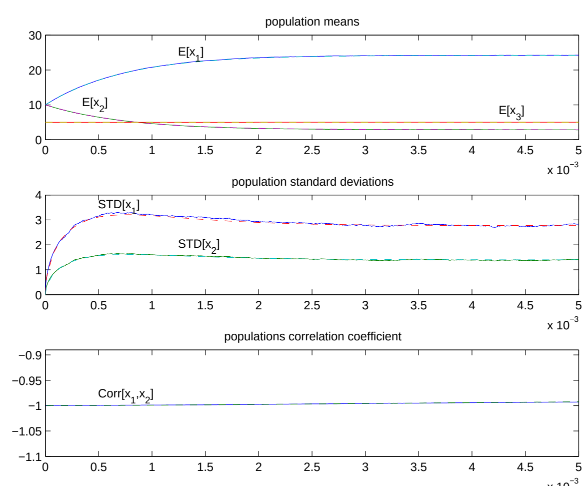
(a) Large population, over a long time scale with initial condition (33)



(b) Large population, over a short time scale with initial condition (34)



(c) Small population, over a long time scale with initial condition (35)



(d) Small population, over short time scale with initial condition (35)

Fig. 11. Comparison between Monte Carlo simulations (solid lines) and the truncated model (29), (32) (dashed lines) for Example 6.

VII. Conclusions and Future Work

We showed that the infinite-dimensional linear moment dynamics of a pSHS can be approximated by a finite-dimensional nonlinear ODE with arbitrary precision. Moreover, we provided a procedure to build this type of approximation. The methodology was illustrated using a varied pool of examples, demonstrating its wide applicability. Several observations arise from these examples, which point to directions for future research:

1) The algorithm proposed to compute the truncation function does not directly take into account the steady-state distribution of the pSHS. This is desirable because for all the examples presented it is difficult to compute the steady-state values of the moments. However, when steady-state distributions are available, they can improve the steady-state accuracy of the truncated system. This was done in Example 2, for which experimental observations indicate that the steady-state distribution of the sending rate appears to be approximately Log-Normal. However, it would be desirable to devise constructions that improve the steady-state accuracy even when this type of insight is not readily available.
2) In all the examples presented, we restricted our attention to truncation functions φ of the form (16) and we only used deterministic distributions to compute φ. Mostly likely, better results could be obtained by considering more general distributions, which may require more general forms for φ.
3) The truncation of pSHSs that model chemical reactions proved especially accurate. This motivates the search for systematic procedures to automatically construct a truncated system from chemical equations such as (6). This issue has been pursued in [18]. Another direction for future research consists of comparing the truncated models obtains here with those in [19].

An additional direction for future research consists of establishing computable bounds on the error between solutions to the infinite-dimensional moments dynamics and to its finite-dimensional truncations.

Appendix

Proof of Theorem 2. For a given $\delta > 0$, let $T > 0$ and $N > 0$ be sufficiently large so that for every solution μ_∞ and ν to (8) and (10) starting at some $t_0 \geq 0$ in Ω_μ and Ω_ν, respectively, we have[4]

$$\beta(\|\mu(t_1) - \nu(t_1)\|, \tau) \leq \frac{\delta}{2}, \qquad \forall \tau \geq T,\ t_1 \geq t_0 \tag{36}$$

$$\frac{(2T)^{N+2}}{(N+2)!}\left(\sup_{\tau \geq t_0}\left\|\frac{\mathrm{d}^{N+2}\mu(\tau)}{\mathrm{d}t^{N+2}}\right\| + \sup_{\tau \geq t_0}\left\|\frac{\mathrm{d}^{N+2}\nu(\tau)}{\mathrm{d}t^{N+2}}\right\|\right) \leq \frac{\delta}{2}. \tag{37}$$

The existence of T is guaranteed by the fact that $\beta \in \mathcal{KL}$ and by the boundedness of ν and μ, as per Assumption 1. For the chosen T, the existence of N is guaranteed by Assumption 1 and the fact that $\lim_{N\to\infty} \frac{(2T)^{N+2}}{(N+2)!} = 0$. The justification for these definitions will become clear shortly.

Let μ_∞ and ν be solutions to (8) and (10) starting in Ω_μ and Ω_ν, respectively, at some time $t_0 \geq 0$. For a given $t_1 \geq t_0$, since Ω_ν is forward invariant, $\nu_1 := \nu(t_1) \in \Omega_\nu$ and therefore, because of Assumption 2, there exists some

[4]The left-hand-sides of (36) and (37) need to add up to δ but other options are possible.

$\hat{\mu}_\infty(t_1) \in \Omega_\mu$ whose first k elements match ν_1 and for which (11) holds. Therefore

$$\begin{aligned}\|\mu(t)-\nu(t)\| &\leq \|\mu(t)-\hat{\mu}(t)\| + \|\hat{\mu}(t)-\nu(t)\| \\ &\leq \beta(\|\mu(t_1)-\nu(t_1)\|, t-t_1) + \|\hat{\mu}(t)-\nu(t)\|, \qquad \forall t \geq t_1.\end{aligned} \tag{38}$$

We now show by induction on i that

$$\frac{\mathrm{d}^i\hat{\mu}(t_1)}{\mathrm{d}t^i} = \frac{\mathrm{d}^i\nu(t_1)}{\mathrm{d}t^i}, \qquad \forall i \in \{0,1,\dots,N+1\}. \tag{39}$$

The base of induction ($i=0$) holds trivially by the way $\hat{\mu}(t_1)$ was selected. Suppose now that the statement hold for some $i \leq N$. Taking $i+1$ derivatives of $\hat{\mu}(t)-\nu(t)$ with respect to t, one obtains

$$\begin{aligned}\frac{\mathrm{d}^{i+1}\big(\hat{\mu}(t)-\nu(t)\big)}{\mathrm{d}t^{i+1}} &= \frac{\mathrm{d}^i}{\mathrm{d}t^i}\Big(\dot{\hat{\mu}}(t)-\dot{\nu}(t)\Big) \\ &= \frac{\mathrm{d}^i}{\mathrm{d}t^i}\Big(A(t)\big(\hat{\mu}(t)-\nu(t)\big)\Big) + \frac{\mathrm{d}^i}{\mathrm{d}t^i}\Big(B(t)\big(C\hat{\mu}_\infty(t)-\bar{\nu}(t)\big)\Big) \\ &= \sum_{j=0}^{i}\binom{i}{j}\frac{\mathrm{d}^{i-j}A(t)}{\mathrm{d}t^{i-j}}\frac{\mathrm{d}^j\big(\hat{\mu}(t)-\nu(t)\big)}{\mathrm{d}t^j} + \sum_{j=0}^{i}\binom{i}{j}\frac{\mathrm{d}^{i-j}B(t)}{\mathrm{d}t^{i-j}}\frac{\mathrm{d}^j\big(C\hat{\mu}_\infty(t)-\bar{\nu}(t)\big)}{\mathrm{d}t^j}.\end{aligned} \tag{40}$$

All the terms $\frac{\mathrm{d}^j\left(\hat{\mu}(t)-\nu(t)\right)}{\mathrm{d}t^j}$ in the first summation are equal to zero at $t=t_1$ because of the induction hypothesis. Moreover, from (12) and (13) we conclude that

$$\frac{\mathrm{d}^jC\hat{\mu}_\infty(t_1)}{\mathrm{d}t^j} = C^j(t_1)\hat{\mu}_\infty(t_1) = (L^j\varphi)(\hat{\mu}(t_1),t_1) = (L^j\varphi)(\nu(t_1),t_1) = \frac{\mathrm{d}^j\bar{\nu}(t_1)}{\mathrm{d}t^j}.$$

This means that all the terms $\frac{\mathrm{d}^j\left(C\hat{\mu}_\infty(t)-\bar{\nu}(t)\right)}{\mathrm{d}t^j}$ in (40) are also equal to zero at $t=t_1$, which finishes the proof of (39).

Expanding $\hat{\mu}(t)$ and $\nu(t)$ as an Nth-order Taylor series around the point $t_1 \geq 0$ and using (39) and (37), conclude that $\|\hat{\mu}(t)-\nu(t)\| \leq \frac{\delta}{2}$, $\forall t \in [t_1, t_1+2T]$. From this and (38), we obtain

$$\|\mu(t)-\nu(t)\| \leq \beta(\|\mu(t_1)-\nu(t_1)\|, t-t_1) + \frac{\delta}{2}, \qquad \forall t \in [t_1, t_1+2T]. \tag{41}$$

We show next by induction that for every integer $i \geq 1$,

$$\|\mu(t)-\nu(t)\| \leq \delta, \qquad \forall t \in [t_0+iT, t_0+(i+1)T]. \tag{42}$$

The basis of induction ($i=1$) is a direct consequence of (41) with $t_1 := t_0$ and (36). Assuming now that (42) holds for some $i \geq 1$, we conclude from (41) with $t_1 := t_0+iT$ that

$$\|\mu(t)-\nu(t)\| \leq \beta(\|\mu(t_0+iT)-\nu(t_0+iT)\|, t-t_0-iT) + \frac{\delta}{2} \leq \delta,$$

$\forall t \in [t_0+(i+1)T, t_0+(i+2)T]$, which finishes the proof of (42). From (41) with $t_1 := t_0$ and (42) if follows that

$$\|\mu(t)-\nu(t)\| \leq \beta(\|\mu(t_0)-\nu(t_0)\|, t-t_0) + \delta, \qquad \forall t \geq t_0,$$

which completes the proof. ■

Proof of Lemma 1. Take a vector $\mu_\infty \in \Omega_\mu$ corresponding to the distribution $P(\cdot\,;q,x)$ and let $\mu_{q_i}^{(m_i)}$ be the moment corresponding to the ith row of μ_∞. Considering a particular row of C that has a one at the ℓth columns and zero at every other column, the corresponding row in the vector equality (17a) can be written as

$$\mu_{q_\ell}^{(m_\ell)} = \lambda_\ell \prod_{i=1}^{k} (\mu_{q_i}^{(m_i)})^{\gamma_{\ell,i}} \quad \Leftrightarrow \quad \begin{cases} x^{(m_\ell)} & q_\ell = q \\ 0 & q_\ell \neq q \end{cases} = \lambda_\ell \prod_{i=1}^{k} \begin{cases} 1 & \gamma_{\ell,i} = 0 \\ x^{(\gamma_{\ell,i} m_i)} & \gamma_{\ell,i} \neq 0,\ q_i = q \\ 0 & \gamma_{\ell,i} \neq 0,\ q_i \neq q \end{cases}$$

where $\lambda_\ell \neq 0$ is the diagonal entry of Λ that corresponds this particular row of C. For $q_\ell = q$ and $x \neq 0$ the left-hand-side is nonzero and therefore we must have $\gamma_{\ell,i} = 0$ whenever $q_i \neq q$, because otherwise the right-hand-side would be zero, which proves 2. Moreover, we must also have

$$x^{(m_\ell)} = \lambda_\ell x^{\left(\sum_{i=1, q_i = q_\ell}^{k} m_i \gamma_{\ell,i} \right)}, \quad \forall x \in \Omega_x,$$

which proves 1 and 3. Consider now the corresponding row in the vector equality (17b), which can be written as

$$\sum_{i=1}^{\infty} a_{\ell,i} \mu_{q_i}^{(m_i)} = \mu_{q_\ell}^{(m_\ell)} \sum_{j=1}^{k} \frac{\gamma_{\ell,j}}{\mu_{q_j}^{(m_j)}} \sum_{i=1}^{\infty} a_{j,i} \mu_{q_i}^{(m_i)} \tag{43}$$

For $q_\ell = q$, since $\gamma_{\ell,j} = 0$ whenever $q_j \neq q = q_\ell$ (because of 2), we conclude that

$$\sum_{\substack{i=1 \\ q_i = q}}^{\infty} a_{\ell,i}\, x^{(m_i)} = \sum_{\substack{j=1 \\ q_j = q}}^{k} \gamma_{\ell,j} \sum_{\substack{i=1 \\ q_i = q}}^{\infty} a_{j,i}\, x^{(m_\ell - m_j + m_i)}, \qquad \forall x \in \Omega_x,$$

which proves 4. For $q_\ell \neq q$, the right-hand-side of (43) is zero and therefore $\sum_{\substack{i=1 \\ q_i = q}}^{\infty} a_{\ell,i}\, x^{(m_i)} = 0$, $\forall x \in \Omega_x$ which proves 5. ∎

REFERENCES

[1] J. P. Hespanha, "Stochastic hybrid systems: Applications to communication networks," in *Hybrid Systems: Computation and Control* (R. Alur and G. J. Pappas, eds.), no. 2993 in Lect. Notes in Comput. Science, pp. 387–401, Berlin: Springer-Verlag, Mar. 2004.

[2] M. H. A. Davis, *Markov models and optimization*. Monographs on statistics and applied probability, London, UK: Chapman & Hall, 1993.

[3] M. L. Bujorianu, "Extended stochastic hybrid systems and their reachability problem," in *Hybrid Systems: Computation and Control*, Lect. Notes in Comput. Science, pp. 234–249, Berlin: Springer-Verlag, Mar. 2004.

[4] J. Hu, J. Lygeros, and S. Sastry, "Towards a theory of stochastic hybrid systems," in *Hybrid Systems: Computation and Control* (N. A. Lynch and B. H. Krogh, eds.), vol. 1790 of *Lect. Notes in Comput. Science*, pp. 160–173, Springer, 2000.

[5] G. Pola, M. L. Bujorianu, J. Lygeros, and M. D. D. Benedetto, "Stochastic hybrid models: An overview," in *Proc. of the IFAC Conf. on Anal. and Design of Hybrid Syst.*, June 2003.

[6] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Series in Electrical Engineering, New York: McGraw-Hill, 3rd ed., 1991.

[7] A. R. Teel, L. Moreau, and D. Nešić, “A unified framework for input-to-state stability in systems with two time scales,” *IEEE Trans. on Automat. Contr.*, vol. 48, pp. 1526–1544, Sept. 2003.

[8] S. Bohacek, J. P. Hespanha, J. Lee, and K. Obraczka, “A hybrid systems modeling framework for fast and accurate simulation of data communication networks,” in *Proc. of the ACM Int. Conf. on Measurements and Modeling of Computer Systems (SIGMETRICS)*, June 2003.

[9] Y. Xu and J. P. Hespanha, “Communication logics for networked control systems,” in *Proc. of the 2004 Amer. Contr. Conf.*, pp. 572–577, June 2004.

[10] Y. Xu and J. P. Hespanha, “Optimal communication logics for networked control systems,” in *Proc. of the 43rd Conf. on Decision and Contr.*, Dec. 2004.

[11] D. T. Gillespie, “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions,” *J. of Computational Physics*, vol. 22, pp. 403–434, 1976.

[12] D. T. Gillespie and L. R. Petzold, “Improved leap-size selection for accelerated stochastic simulation,” *J. of Chemical Physics*, vol. 119, pp. 8229–8234, Oct. 2003.

[13] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, “Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method,” *J. of Chemical Physics*, vol. 119, pp. 12784–12794, Dec. 2003.

[14] J. P. Hespanha, “Polynomial stochastic hybrid systems,” in *Hybrid Systems: Computation and Control* (M. Morari and L. Thiele, eds.), no. 3414 in Lect. Notes in Comput. Science, pp. 322–338, Berlin: Springer-Verlag, Mar. 2005.

[15] J. P. Hespanha, “A model for stochastic hybrid systems with application to communication networks,” *Nonlinear Analysis*, Special Issue on Hybrid Systems, vol. 62, pp. 1353–1383, Sept. 2005.

[16] G. Irlam, “Unix file size survey – 1993.” Available at `http://www.base.com/gordoni/ufs93.html`, Nov. 1994.

[17] A. Singh and J. P. Hespanha, “Models for multi-species chemical reactions using polynomial stochastic hybrid systems,” in *Proc. of the 44th Conf. on Decision and Contr.*, Dec. 2005.

[18] J. P. Hespanha and A. Singh, “Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems,” *Int. J. on Robust Control*, Special Issue on Control at Small Scales: Issue 1, vol. 15, pp. 669–689, Sept. 2005.

[19] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier Science, 2001.