

# Moment Propagation of Discrete-Time Stochastic Polynomial Systems using Truncated Carleman Linearization

Sasinee Pruekprasert \* Toru Takisaka \* Clovis Eberhart \*\*,\*\*\*  
Ahmet Cetinkaya \* Jérémy Dubut \*\*,\*\*\*

\* National Institute of Informatics, Tokyo, Japan  
(e-mail: {sasinee, cetinkaya, eberhart, takisaka, dubut}@nii.ac.jp).

\*\* Japanese-French Laboratory of Informatics, UMI 3527, CNRS,  
Tokyo, Japan

---

**Abstract:** We propose a method to compute an approximation of the moments of a discrete-time stochastic polynomial system. We use the Carleman linearization technique to transform this finite-dimensional polynomial system into an infinite-dimensional linear one. After taking expectation and truncating the induced deterministic dynamics, we obtain a finite-dimensional linear deterministic system, which we then use to iteratively compute approximations of the moments of the original polynomial system at different time steps. We provide upper bounds on the approximation error for each moment and show that, for large enough truncation limits, the proposed method precisely computes moments for sufficiently small degrees and numbers of time steps. We use our proposed method for safety analysis to compute bounds on the probability of the system state being outside a given safety region. Finally, we illustrate our results on two concrete examples, a stochastic logistic map and a vehicle dynamics under stochastic disturbance.

*Keywords:* Stochastic systems, nonlinear systems, probabilistic safety analysis, moment computation, Carleman linearization

---

## 1. INTRODUCTION

Ensuring safety in cyber-physical systems under uncertainties is an important and challenging task. For instance, automated driving systems are required to take into account the uncertainties in the environment and those in control operations to avoid collision with obstacles (Bry and Roy, 2011). Motion planning algorithms in automated driving typically compute probabilities of collision in different potential paths of a vehicle by identifying possible future states of the vehicle under the effects of measurement and actuation uncertainties.

To quantify how uncertainties affect the state of the system over time, one often uses the idea of *mean and covariance propagation*. Particularly, if the dynamical system is linear and subject to additive Gaussian noise, then this state follows a Gaussian distribution, and the evolution of its mean and covariance can be described by a linear equation. Given the initial mean and covariance, one can then use this linear equation to compute mean and covariance iteratively at each future time instant. This idea has been used in conjunction with a Kalman-filtering approach in motion planning by Bry and Roy (2011) and Banzhaf et al. (2018). Here, we are interested in whether similar iterative techniques can be used to compute mean, covariance, and even higher moments when the dynamical system is

nonlinear and the noise is not necessarily additive and Gaussian. We provide a positive answer to this question.

In this paper, we propose a method to compute approximations of the moments through truncated Carleman linearization (Steeb and Wilhelm, 1980). Specifically, we consider a discrete-time nonlinear stochastic polynomial system with a model that allows us to describe additive and multiplicative stochastic noise in a unified manner through randomly varying coefficient matrices. We use the Carleman linearization technique on this system to obtain an infinite-dimensional linear one that characterizes the evolution of all Kronecker powers of the state variables. We observe that the expected dynamics of this linear system relates the moments of the random coefficient matrices to the moments of the state. By using a truncation approach, we then obtain a finite-dimensional linear system, which we use to compute an approximation of the mean and covariance of the system state and its higher moments up to a given truncation limit.

We then propose a framework to verify the stochastic safety of the system via tail probability estimation techniques (Gray and Wang, 1991). This framework relies on the efficient computation of error bounds for our moment approximations, which we provide. The exact error is the sum of many terms whose number grows quickly with respect to the dimension of the system and time, and we give an efficient way to compute a bound by splitting the sum into two parts: the part (e.g., of linear or polynomial

---

\* The authors are supported by ERATO HASUO Metamathematics for Systems Design Project No. JPMJER1603, JST; J. Dubut is also supported by Grant-in-aid No. 19K20215, JSPS.

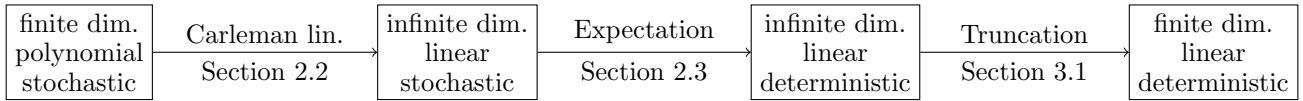


Fig. 1. The different steps of the method

size in the dimension of the system) of the sum that strongly contribute to the error and on which we make a fine computation, and the rest on which we do a cruder but faster computation.

We demonstrate the utility of our approach with two case studies. First, we compute moments for the scalar stochastic logistic map with uniformly distributed random growth/decay rates. We then look at a higher-dimensional polynomial discrete-time vehicle model obtained via second-order Taylor expansion of a kinematic bicycle model. We use our approach to investigate how the uncertainty in the acceleration affects the vehicle's future positions. In particular, our approach uses the moments of the initial state of the vehicle, together with the moments of its noisy acceleration. The moments of the initial state of the vehicle is obtained by using a noisy measurement of the state and the information on the statistical properties of the measurement error. Our framework allows moment computation in a receding horizon fashion. Specifically, at each time step, new measurements can be used for specifying the moments of the new "initial" state, which can then be used to compute the moments associated with time instants that are further in the future.

Carleman linearization is a well-known approach in nonlinear dynamical systems literature. It has been used in several pieces of work for approximation (Bellman and Richardson, 1963; Steeb and Wilhelm, 1980; Al-Tuwaim et al., 1998), and more recently for optimal control (Amini et al., 2019) and model predictive control of continuous-time deterministic nonlinear systems (Hashemian and Armaou, 2019). Results on Carleman linearization for stochastic systems are relatively scarce. For Ito-type stochastic differential equations with bilinear noise, Wong (1983) used Carleman linearization to obtain differential equations describing the evolution of the moments. Moreover, Rauh et al. (2009) used a series-expansion approach together with the Carleman linearization technique to approximate a continuous-time nonlinear stochastic system with a discrete-time linear system under additive noise. They then use the linear system for state estimation via a Kalman filter. Recently, Cacace et al. (2014, 2017) developed Carleman linearization-based filters for nonlinear continuous-time stochastic systems with Wiener noise and periodic measurements. Although the discrete-time representation of the sampled-data filter equations in those pieces of work involves multiplicative noise, the models and the problem formulations there differ from those in our paper: on the one hand, they can describe non-polynomial dynamics, but on the other hand, they can only handle Wiener noises. We note that our truncation error analysis is partly motivated by the analysis in Forets and Pouly (2017), which provides tight approximation error bounds for the truncated Carleman linearization approach in deterministic continuous-time systems. We also note that in addition to Carleman linearization, Koopman operators are also used for deriving infinite-dimensional

linear equations to describe finite-dimensional nonlinear systems Goswami and Paley (2017); Mesbahi et al. (2019), although the techniques for analysis of these two methods are quite different.

The rest of the paper is organized as follows (see also Figure 1 for the specific subsections where the main steps of the method are presented). In Section 2, we start from a finite-dimensional stochastic polynomial system, on which we apply the Carleman linearization technique to get an infinite-dimensional linear stochastic system, which we then turn into a deterministic one by taking expectation. We present our truncation-based moment approximation method and study the error it introduces in Section 3, in which we also give an application of this bound. We provide two numerical examples to demonstrate our approach in Section 4. Finally, in Section 5, we conclude the paper.

**Notations.** We use  $\mathbb{N}$  for the set of non-negative integers. We denote by  $M_1 \otimes M_2$  the Kronecker product of two matrices  $M_1$  and  $M_2$ . Moreover, we use  $M^{[k]}$  to denote the  $k$ th Kronecker power of the matrix  $M$ , given by  $M^{[0]} = 1$  and  $M^{[k]} = M^{[k-1]} \otimes M$  for  $k > 0$ . The dimensions of the matrices we use grow rapidly, so we will use  $n^{\leq k}$  as a shorthand for  $\sum_{i=0}^k n^i$  when manipulating dimensions of vectors or matrices.

## 2. CARLEMAN LINEARIZATION FOR STOCHASTIC POLYNOMIAL SYSTEMS

In this section, we use the Carleman linearization technique to turn a finite-dimensional discrete-time stochastic polynomial system into an infinite-dimensional one, then take expectation to get a deterministic system that describes the evolution of all moments of the system state.

### 2.1 Discrete-Time Stochastic Polynomial Systems

We focus on computing the moments of the state of the following finite-dimensional discrete-time stochastic polynomial system

$$\begin{aligned} x(t+1) &= \sum_{i=0}^{d_S} F_i(t)x^{[i]}(t), \quad t \in \mathbb{N}, \\ x(0) &= x_0, \end{aligned} \quad (1)$$

where  $x(t) \in \mathbb{R}^n$  is the state vector and  $F_i(t) \in \mathbb{R}^{n \times n^i}$ ,  $i \in \{0, 1, \dots, d_S\}$ , are randomly-varying coefficient matrices. To be more precise, we assume given a probabilistic space  $\Omega$ , then  $x_0$  (resp.  $x(t)$ ,  $F_i(t)$ ) is actually a measurable function from  $\Omega$  to  $\mathbb{R}^n$  (resp.  $\mathbb{R}^n$ ,  $\mathbb{R}^{n \times n^i}$ ). We will omit to mention  $\Omega$  in this paper.

We remark that (1) is useful for modeling dynamics with both additive and multiplicative noise terms: the vector  $F_0(t) \in \mathbb{R}^n$  represents additive noise, as  $x^{[0]} = 1$ , while the terms  $F_1(t), \dots, F_{d_S}(t)$  characterize the effects of multiplicative noise.

### 2.2 Carleman Linearization

We use Carleman linearization to obtain an infinite-dimensional linear system that describes the evolution of the Kronecker powers of the state  $x(t)$ . By defining

$$F(t) \triangleq [F_0(t) \ F_1(t) \ F_2(t) \ \dots \ F_{d_S}(t)] \in \mathbb{R}^{n \times n^{\leq d_S}}$$

and

$$y(k, t) \triangleq [1 \ x(t)^\top \ x^{[2]}(t)^\top \ \dots \ x^{[k]}(t)^\top]^\top \in \mathbb{R}^{n^{\leq k}}, \quad (2)$$

the dynamical system (1) can be rewritten as

$$x(t+1) = F(t)y(d_S, t).$$

Therefore, for all  $j \in \mathbb{N}$ , we have

$$x^{[j]}(t+1) = (F(t)y(d_S, t))^{[j]}.$$

By using the *mixed-product property* of the Kronecker product (i.e.,  $(A \otimes B)(C \otimes D) = AC \otimes BD$ ; see Section 13.2 of Laub (2005)), we obtain

$$\begin{aligned} x^{[j]}(t+1) &= F^{[j]}(t)y^{[j]}(d_S, t) \\ &= \sum_{k=0}^{j d_S} \left( \sum_{(i_l)_{l \leq j} \in H_{j,k}} F_{i_1}(t) \otimes \dots \otimes F_{i_j}(t) \right) x^{[k]}(t), \end{aligned}$$

for  $j \in \mathbb{N}$ , where

$$H_{j,k} \triangleq \left\{ (i_l)_{l \leq j} \mid \sum_{l=1}^j i_l = k \text{ and } i_l \leq d_S \right\}.$$

We thus get the infinite-dimensional linear system

$$\begin{aligned} x^{[j]}(t+1) &= \sum_{k=0}^{j d_S} \left( \sum_{(i_l)_{l \leq j} \in H_{j,k}} \bigotimes_{m=1}^j F_{i_m}(t) \right) x^{[k]}(t), \quad (3) \\ x^{[j]}(0) &= x_0^{[j]}, \end{aligned}$$

which describes the evolution of all Kronecker powers  $x^{[0]}(t), x^{[1]}(t), \dots$  of the state  $x(t)$ .

In order to give a simpler description of the system, we introduce the matrix  $A_{j,k}(t) \in \mathbb{R}^{n^j \times n^k}$  given by

$$A_{j,k}(t) \triangleq \sum_{(i_l)_{l \leq j} \in H_{j,k}} F_{i_1}(t) \otimes \dots \otimes F_{i_j}(t).$$

Note in particular that  $A_{0,0}(t) = 1$  since there is only the empty sequence in  $H_{0,0}$ , and  $A_{j,k}(t) = 0$  if  $j = 0$  and  $k > 0$  or  $k > j d_S$  since  $H_{j,k}$  is then empty. We also introduce, for all  $N, M \in \mathbb{N}$ , the matrix  $A(t; N, M)$  defined by blocks as  $A(t; N, M) = [A_{j,k}(t)]_{\substack{j \leq N \\ k \leq M}}$ , that is,

$$A(t; N, M) \triangleq \begin{bmatrix} A_{0,0}(t) & \dots & A_{0,M}(t) \\ \vdots & \ddots & \vdots \\ A_{N,0}(t) & \dots & A_{N,M}(t) \end{bmatrix} \in \mathbb{R}^{n^{\leq N} \times n^{\leq M}}.$$

Then, for any  $k \in \mathbb{N}$ , we have

$$\begin{aligned} y(k, t+1) &= [1 \ x(t+1)^\top \ x^{[2]}(t+1)^\top \ \dots \ x^{[k]}(t+1)^\top]^\top \\ &= A(t; k, k d_S) [1 \ x(t)^\top \ x^{[2]}(t)^\top \ \dots \ x^{[k d_S]}(t)^\top]^\top \\ &= A(t; k, k d_S)y(k d_S, t). \end{aligned} \quad (4)$$

### 2.3 Moment Equations

We now derive the deterministic system that describes the evolution of the moments of  $x(t)$  by taking expectation in (4). This gives

$$\mathbb{E}[y(k, t+1)] = \mathbb{E}[A(t; k, k d_S)y(k d_S, t)].$$

We make the following assumptions concerning the coefficient matrices and the random initial state  $x_0$ .

*Assumption 1.* The matrices  $\{F(t) \mid t \in \mathbb{N}\}$  are independent. Moreover,  $x_0$  is independent of  $F(t)$ ,  $t \in \mathbb{N}$ .

*Assumption 2.* The matrices  $\{F(t) \mid t \in \mathbb{N}\}$  are identically distributed.

Note that Assumptions 1 and 2 are not overly restrictive and they hold in fairly general situations as we discuss in Section 4. Notice also that under Assumption 1, matrices  $F_i(t)$  and  $F_j(t)$  are still allowed to statistically depend on each other. Moreover, Assumption 2 allows us to obtain a “time-invariant” method to compute moments, further yielding computational advantage.

By iteration of (4), we get that  $y(k d_S, t)$  is given by

$$\begin{aligned} y(k d_S, t) &= A(t-1; k d_S, k d_S^2) \dots A(0; k d_S^t, k d_S^{t+1}) \\ &\quad \cdot y(k d_S^{t+1}, 0), \quad t \in \mathbb{N}. \end{aligned} \quad (5)$$

It follows from Assumption 1 that  $A(t; k, k d_S)$  and  $y(k d_S, t)$  in (4) are mutually independent. To see this, observe that  $A(t; k, k d_S)$  is composed of the matrices  $F_i(t)$ , which are independent of  $x_0$  and  $F(t-1), \dots, F(0)$ , which determine  $y(k d_S, t)$  as given by (5).

It then follows that

$$\mathbb{E}[y(k, t+1)] = \mathbb{E}[A(t; k, k d_S)]\mathbb{E}[y(k d_S, t)].$$

Here again, to give a simpler description of the system, we introduce new matrices. Notice that the coefficients of the matrix  $\mathbb{E}[A(t; N, M)]$  are products of the moments of coefficients of  $F(t)$  and thus independent of  $t$  by Assumption 2.

We denote this matrix by  $E(N, M) \in \mathbb{R}^{n^{\leq N} \times n^{\leq M}}$ , emphasizing the fact that it is independent of the time. Similarly, we denote by  $E_{i,j} \in \mathbb{R}^{n^i \times n^j}$  the matrix  $\mathbb{E}[A_{i,j}(t)]$ . The equation above can thus be rewritten to give the following system

$$\begin{aligned} \mathbb{E}[x^{[j]}(t+1)] &= \sum_{k=0}^{j d_S} E_{j,k} \mathbb{E}[x^{[k]}(t)], \quad t \in \mathbb{N}, \\ \mathbb{E}[x^{[j]}(0)] &= \mathbb{E}[x_0^{[j]}]. \end{aligned} \quad (6)$$

Since the matrices  $E_{j,k}$  only depend on the moments of the matrices  $F(\cdot)$ , they can be computed offline.

### 3. MOMENT APPROXIMATION THROUGH TRUNCATION

We will now introduce an approach that allows us to compute *approximations* of the moments of  $x(t)$ . This truncation approach is critical, as an exact computation of the moments is impossible from a practical point of view. Indeed, by (2), computing the first  $k$  moments at time  $t$  amounts to computing  $\mathbb{E}[y(k, t)]$ . So, by iteration of (6),

$$\mathbb{E}[y(k, t)] = E(k, kd_S) \cdots E(kd_S^{t-1}, kd_S^t) \mathbb{E}[y(kd_S^t, 0)].$$

As this indicates, exact computation of the first  $k$  moments requires the knowledge of matrices  $E(k, kd_S)$ ,  $E(kd_S, kd_S^2)$ ,  $\dots$ ,  $E(kd_S^{t-1}, kd_S^t)$  of exponentially increasing dimensions, making any practical computation unrealistic. We thus resort to a truncation approach where we fix a truncation limit and consider a matrix of fixed size to approximately compute the moments of  $x(t)$ .

### 3.1 Approximate Moments and the Truncated System

In this section, we define the system that we use to compute approximations of the moments of  $x(t)$ .

We fix  $N_T \in \mathbb{N}$  and define  $\tilde{x}_i(t) \in \mathbb{R}^{n_i}$ ,  $i \in 1, \dots, N_T$ , by

$$\begin{aligned} & [1 \ \tilde{x}_1(t)^\top \ \cdots \ \tilde{x}_{N_T}(t)^\top]^\top \\ &= E(N_T, N_T)^t [1 \ \mathbb{E}[x_0]^\top \ \cdots \ \mathbb{E}[x_0^{[N_T]}]^\top]^\top. \end{aligned} \quad (7)$$

Here, the vector  $\tilde{x}_i(t)$  represents an approximation of the moment  $\mathbb{E}[x^{[i]}(t)]$  that is computed using only our knowledge of the first  $N_T$  moments of  $x_0$ .

By letting  $\tilde{y}(t) \triangleq [1 \ \tilde{x}_1^\top(t) \ \tilde{x}_2^\top(t) \ \cdots \ \tilde{x}_{N_T}^\top(t)]^\top$ , we obtain the so-called ‘‘truncated system’’, which is a discrete-time linear time-invariant system given by

$$\begin{aligned} \tilde{y}(t+1) &= E(N_T, N_T) \tilde{y}(t), \quad t \in \mathbb{N}, \\ \tilde{y}(0) &= [1 \ \mathbb{E}[x_0(t)]^\top \ \cdots \ \mathbb{E}[x_0^{[N_T]}(t)]^\top]^\top. \end{aligned} \quad (8)$$

The truncated system allows us to iteratively compute approximations of the moments of  $x(t)$  at consecutive time instants. Moreover, the approach only requires an offline computation of the matrix  $E(N_T, N_T)$ .

### 3.2 Approximation Error Bounds

We now investigate the approximation error introduced by truncation. Let  $e_i(t) \in \mathbb{R}^{n_i}$  denote the approximation error of the  $i$ th moment, that is,

$$e_i(t) \triangleq \mathbb{E}[x^{[i]}(t)] - \tilde{x}_i(t).$$

In what follows, we provide upper bounds to  $\|e_i(t)\|$  (the error on the  $i$ th moment introduced by the approximation) for any norm  $\|\cdot\|$  induced by an inner product. These bounds allow us to use various techniques to study the distribution of the state at future time steps. We illustrate this in Section 3.3 by using tail probability approximations (Gray and Wang, 1991) to compute an upper bound of the probability to be outside of a given safe region.

Let  $j_0 \in \{0, \dots, N_T\}$ . Our goal is to obtain an upper bound to  $\|e_{j_0}(t)\|$ . First, by (6), we have

$$\begin{aligned} \mathbb{E}[x^{[j_0]}(t)] &= \sum_{j_1=0}^{j_0 d_S} E_{j_0, j_1} \mathbb{E}[x^{[j_1]}(t-1)] \\ &= \sum_{j_1=0}^{j_0 d_S} E_{j_0, j_1} \sum_{j_2=0}^{j_1 d_S} E_{j_1, j_2} \mathbb{E}[x^{[j_2]}(t-2)] \\ &= \sum_{j_1=0}^{j_0 d_S} E_{j_0, j_1} \sum_{j_2=0}^{j_1 d_S} \cdots \sum_{j_t=0}^{j_{t-1} d_S} E_{j_{t-1}, j_t} \mathbb{E}[x_0^{[j_t]}], \end{aligned}$$

and similarly, by (7),

$$\tilde{x}_{j_0}(t) = \sum_{j_1=0}^{N_T} E_{j_0, j_1} \sum_{j_2=0}^{N_T} E_{j_1, j_2} \cdots \sum_{j_t=0}^{N_T} E_{j_{t-1}, j_t} \mathbb{E}[x_0^{[j_t]}].$$

By observing that  $E_{j,k} = 0$  if  $k > jd_S$ , we obtain from the two equations above,

$$\begin{aligned} e_{j_0}(t) &= \mathbb{E}[x^{[j_0]}(t)] - \tilde{x}_{j_0}(t) \\ &= \sum_{j_1=N_T+1}^{j_0 d_S} E_{j_0, j_1} \sum_{j_2=0}^{j_1 d_S} \cdots \sum_{j_t=0}^{j_{t-1} d_S} E_{j_{t-1}, j_t} \mathbb{E}[x_0^{[j_t]}] \\ &\quad + \sum_{j_1=0}^{N_T} E_{j_0, j_1} \sum_{j_2=N_T+1}^{j_1 d_S} \cdots \sum_{j_t=0}^{j_{t-1} d_S} E_{j_{t-1}, j_t} \mathbb{E}[x_0^{[j_t]}] \\ &\quad + \cdots \\ &\quad + \sum_{j_1=0}^{N_T} E_{j_0, j_1} \cdots \sum_{j_{t-1}=0}^{N_T} E_{j_{t-2}, j_{t-1}} \sum_{j_t=N_T+1}^{j_{t-1} d_S} E_{j_{t-1}, j_t} \mathbb{E}[x_0^{[j_t]}]. \end{aligned} \quad (9)$$

As an immediate consequence, we get the following:

**Proposition 3.** Consider the truncated approximation of the moments of system (1) with truncation limit  $N_T \in \mathbb{N}$ . If  $j_0 d_S^t \leq N_T$ , then  $\tilde{x}_{j_0}(t) = \mathbb{E}[x^{[j_0]}(t)]$ .

**Proof.** We show that, if  $j_0 d_S^t \leq N_T$ , then  $\|e_{j_0}(t)\| = 0$ . To this end, it is enough to show that, for all lines of (9) and sequences  $(j_1, \dots, j_t)$  of relevant indices, at least one  $E_{j_{k-1}, j_k}$  is equal to 0. Let us pick the  $i$ th line and any sequence as above. In particular,  $j_{i-1} d_S > N_T$ , hence  $j_{i-1} > j_0 d_S^{t-1} \geq j_0 d_S^{i-1}$ , so there exists  $k \in \{1, \dots, i-1\}$  such that  $j_k > j_{k-1} d_S$ , which proves the desired result.  $\square$

This shows that, for large values of truncation limit  $N_T$ , the proposed method computes *exact* moments  $\mathbb{E}[x^{[j_0]}(t)]$  for  $j_0$  and  $t$  small enough. Note that this is due to the *discrete-time* nature of the finite-dimensional polynomial system (1). In the continuous-time case, approximation errors cannot be avoided in general (Forets and Pouly, 2017).

However, the exact value of  $e_{j_0}(t)$  is generally hard to compute. Indeed, since  $\mathbb{E}[x^{[i]}(t)] = \tilde{x}_i(t) + e_i(t)$ , if  $e_i(t)$  could efficiently be computed, then so would  $\mathbb{E}[x^{[i]}(t)]$ , and there would be no need in using the truncated system. In the rest of this section, we thus come up with several upper bounds for  $\|e_{j_0}(t)\|$  that can be efficiently computed.

First, observe that  $\xi = \max_{0 \leq j \leq j_0 d_S^t} \|\mathbb{E}[x_0^{[j]}]\|$  can be efficiently computed in some cases. An obvious situation is when the position  $x_0$  is determined, in which case we have  $\xi = \max\{1, \|x_0\|^{j_0 d_S^t}\}$ . Another case is when  $x_0$  obeys a well-known distribution whose moments are easy to compute, such as uniform or normal distributions. Nevertheless, another case is when the system satisfies  $x(t) \in \mathbb{R}$  and  $0 \leq x_0 \leq 1$ , in which case  $\|\mathbb{E}[x_0^{[j]}]\|$  is decreasing and we have  $\xi = \|\mathbb{E}[x_0^{[0]}\| = 1$ . Using  $\xi$ , we can derive bounds for  $e_{j_0}(t)$  by first rewriting (9) as

$$e_{j_0}(t) = \sum_{j=0}^{j_0 d_S^t} \tilde{E}_j \mathbb{E}[x_0^{[j]}], \quad (10)$$

where  $\tilde{E}_j \in \mathbb{R}^{n^{j_0} \times n^j}$  is constructed by additions and multiplications of  $E_{i,j}$ , and thus can be computed offline (note that  $\tilde{E}_j$  is dependent on  $t$ , but we keep this implicit for readability). From this, we can derive a global bound

$$\|e_{j_0}(t)\| \leq \xi \sum_{j=0}^{j_0 d_S^t} \|\tilde{E}_j\|,$$

where  $\sum_{j=0}^{j_0 d_S^t} \|\tilde{E}_j\|$  can be computed offline.

We can further refine (10) to consider a single line  $i \leq n^{j_0}$  of this equation, for which we get  $(e_{j_0}(t))_i = \sum_{j=0}^{j_0 d_S^t} v_{j,i} \mathbb{E}[x_0^{[j]}]$ , where  $v_{j,i} \in \mathbb{R}^{1 \times n^j}$  is the  $i$ th row of  $\tilde{E}_j$ . By repeated application of triangle and Cauchy-Schwarz inequalities, we obtain

$$|(e_{j_0}(t))_i| \leq \xi \sum_{j=0}^{j_0 d_S^t} \|v_{j,i}\|,$$

where  $\sum_{j=0}^{j_0 d_S^t} \|v_{j,i}\|$  can also be computed offline.

This bound can, however, be crude in practice as the norm gets distributed over all sums and products. Here we show how to compute tighter bounds while maintaining a reasonable computational cost. For any subset  $J \subseteq \{0, \dots, j_0 d_S^t\}$ , we have the following:

$$|(e_{j_0}(t))_i| \leq \left| \sum_{j \in J} v_{j,i} \mathbb{E}[x_0^{[j]}] \right| + \xi_J \sum_{j \notin J} \|v_{j,i}\|, \quad (11)$$

where  $\xi_J = \max_{j \notin J} \|\mathbb{E}[x_0^{[j]}]\|$ .

The idea is that  $J$  is a set of indices where one should avoid distributing the norm over the sum. One should pick  $J$  to consist of those indices where the distribution is too crude and makes the error bound loose. In order for this method to be computationally efficient, one should pick  $J$  that is of relatively small size, e.g.,  $|J| = \mathcal{O}(t)$ . One possible way to choose  $J$  is to fix a size  $k$  and return the set of  $k$  indices such that  $\|v_{j,i}\|$  is largest for those indices; another way is to return the set of  $k$  indices such that  $\|\mathbb{E}[x_0^{[j]}]\|$  is largest.

*Example 4.* Suppose  $x_0$  is drawn from a truncated normal distribution over the interval  $[0, 1]$ . Then  $\xi = 1$  and  $\xi_J = \|\mathbb{E}[x_0^{[j_{\min}]}]\|$ , where  $j_{\min}$  is the smallest number that is not in  $J$ . In particular, if  $J$  is chosen as the set of indices such that  $\|\mathbb{E}[x_0^{[j_{\min}]}]\|$  is largest among  $j = 0, \dots, j_0 d_S^t$ , then we have  $J = \{0, \dots, |J| - 1\}$  and  $\xi_J = \|\mathbb{E}[x_0^{[|J|]}\|$ .

### 3.3 Tail Probability Approximation

In this section, we use the error bound introduced by the truncated system, which is derived in the previous section, to give a lower bound on the probability of staying inside a given safe region.

Suppose we know some approximations  $(\tilde{x}_2(t))_{ni+j}$  (henceforth denoted  $(\tilde{x}_2(t))_{i,j}$ ) of expectations  $\mathbb{E}[x_i(t)x_j(t)]$  for

all  $i, j \leq n$  (where  $n$  is the dimension of the state space, as in (1)), with respective error bounds  $\varepsilon_{i,j}$ , as well as approximations  $(\tilde{x}_1(t))_i$  of  $\mathbb{E}[x_i(t)]$  with error bounds  $\varepsilon_i$ . Finally, assume that we know a global error bound  $\|\tilde{x}_1(t) - \mathbb{E}[x(t)]\| \leq \varepsilon$ . These different bounds can be found by applying the techniques in the previous section.

We can then bound the probability to be outside of a “safe” region centered on  $\tilde{x}_1(t)$  by the following proposition.

*Proposition 5.* Let  $\|\cdot\|$  be the Euclidean norm. For  $\alpha > \varepsilon$ ,

$$\begin{aligned} \mathbb{P}(\|x(t) - \tilde{x}_1(t)\| \geq \alpha) &\leq \frac{\sum_{i=1}^n (\tilde{x}_2(t))_{i,i} + \varepsilon_{i,i} - \max\{0, (|\tilde{x}_1(t))_i| - \varepsilon_i\}^2}{(\alpha - \varepsilon)^2}. \end{aligned}$$

**Proof.** Since  $\alpha > \varepsilon$ , by Chebyshev’s inequality:

$$\begin{aligned} \mathbb{P}(\|x(t) - \tilde{x}_1(t)\| \geq \alpha) &\leq \mathbb{P}(\|x(t) - \mathbb{E}[x(t)]\| \geq \alpha - \varepsilon) \\ &\leq \frac{\mathbb{E}[\|x(t) - \mathbb{E}[x(t)]\|^2]}{(\alpha - \varepsilon)^2} \\ &= \frac{\sum_{i=1}^n \mathbb{E}[x_i^2(t)] - \mathbb{E}[x_i(t)]^2}{(\alpha - \varepsilon)^2}, \end{aligned}$$

and we conclude using  $\mathbb{E}[x_i^2(t)] \leq (\tilde{x}_2(t))_{i,i} + \varepsilon_{i,i}$  and  $\mathbb{E}[x_i(t)]^2 = |\mathbb{E}[x_i(t)]|^2 \geq \max\{0, (|\tilde{x}_1(t))_i| - \varepsilon_i\}^2$ .  $\square$

The safety region discussed in Proposition 5 corresponds to a ball in  $\mathbb{R}^n$  with radius  $\alpha$ . We note that safety regions can be generalized to ellipsoids by changing the norm  $\|\cdot\|$  to a weighted norm  $\|\cdot\|_P$  induced by a positive-definite matrix  $P \in \mathbb{R}^n$ . Moreover, semi-norms induced by positive semi-definite matrices can be used for problems where the safety of only certain entries of the state vector is explored. A similar argument may be used to prove a general formula for any semi-norm induced by a positive semi-definite matrix, but we omit it as it will not be used in the examples presented in this paper.

## 4. NUMERICAL EXAMPLES

In this section, we provide two numerical examples and experimental results to illustrate our techniques.

### 4.1 Stochastic Logistic Map

Consider the stochastic logistic map as studied by Athreya and Dai (2000), which is given by

$$\begin{aligned} x(t+1) &= r(t)x(t)(1-x(t)), \quad t \in \mathbb{N}, \\ x(0) &= x_0, \end{aligned}$$

where  $x_0, r(0), r(1), \dots$  are mutually independent random variables;  $x_0$  takes values in  $[0, 1]$  and  $r(t)$  all take values in  $[0, 4]$ . The scalar  $x(t) \in [0, 1]$  represents the population of a species subject to growth rate  $r(t)$ . This system can equivalently be represented by (1) with  $d_S = 2$ ,  $F_0(t) = 0$ ,  $F_1(t) = r(t)$ , and  $F_2(t) = -r(t)$ .

For experiments, we chose all  $r(t)$  to be uniformly distributed over the interval  $[0.3, 0.7]$ , and  $x_0$  to follow a normal distribution of mean 0.5 and standard deviation 0.1 truncated to  $[0, 1]$ .

*Moment Approximation via Truncated System.* We first compare our moment approximations for different truncation limits to the true value of the moments (computed

using our method with  $N_T = 256$ ). In Figure 2, we plot the first and second moments of the truncated system with different truncation limits  $N_T$ . In most cases, we observe that the system with the higher truncation limit gives a better approximation. There is one exception in the second moment approximation at  $t = 6$ ; the system with the lower limit gives a better approximation. This behavior seems to be similar to the one in Forests and Pouly (2017), where a truncated system with a higher  $N_T$  diverges to infinity more quickly, while it gives a better approximation over a longer time interval. (In this case, this is due to the fact that the actual system is converging to 0, and a system with truncation limit  $N_T$  diverges to infinity as a polynomial of degree  $N_T$ .) We also observe that larger  $N_T$  is required to obtain good approximations of higher moments. This is a natural consequence, as truncation discards more information on the dynamics of higher moments.

*Error Bound on Moment Approximations.* Next, we evaluate our approximation method of error bounds for moments. Figure 3 shows error bounds given by the inequality (11) with different sizes of  $J$ , with parameters  $N_T = 16$ ,  $t = 4$ , and  $j_0 = 2$ . The index set  $J$  contains the indices  $j$  where  $\|E[x_0^{[j]}\|$  is the largest, i.e.,  $J = \{0, \dots, |J| - 1\}$ . We observe that the error bound quickly decreases as  $|J|$  increases. This supports our expectation that we can use our error bounds for tail probability analysis with larger parameters ( $t$ ,  $N_T$ , and  $d_S$ ).

*Tail Probability Analysis.* Lastly, we provide a result on tail probability analysis via the method in Section 3.3. We computed the error bound for  $0 \leq t \leq 5$  using (11) with  $N_T = 16$  and  $|J| = 6t$ , where  $J$  is taken in the same way as the previous section. The results of the analysis are depicted in Figure 4. Orange intervals indicate the 95%-probability neighborhoods of  $\tilde{x}_1(t)$  computed by using Proposition 5. Blue intervals with a solid line indicate the region where 95% of 10000 Monte Carlo simulations closest to its mean belong. Dotted intervals indicate the range of 10000 Monte Carlo simulations.

We observe that the size of safety intervals given by our tail probability analysis is reasonably small. It becomes cruder in later time steps, especially at  $t = 5$ . This is expected, as the approximation error of moments, which is a bottleneck in refining the error bounds, becomes larger as time progresses (cf. approximate 2nd moment in Fig. 2).

There are two major advantages of our method compared to Monte Carlo simulation. One is that our technique computes moment approximations much faster than Monte Carlo (even for large  $N_T$  and a small number of samples) because we do not rely on generating random numbers. This advantage is highlighted in Table 1, which contains the online computation times for Monte Carlo simulations and our approach, averaged over 100 runs. Another advantage is that our safety interval gives a theoretical guarantee on probabilistic safety that cannot be achieved by Monte Carlo simulations.

#### 4.2 Vehicle Dynamics

Our second example is an application to autonomous driving. For safety guarantees, autonomous vehicles need to

Table 1. Comparison of online computation times.

Method	Monte Carlo		Moment propagation			
	num. samples		$N_T$			
	10	$10^4$	4	16	64	256
Time ( $\mu s$ )	$2.9e10^3$	$3.4e10^6$	11	14	30	93

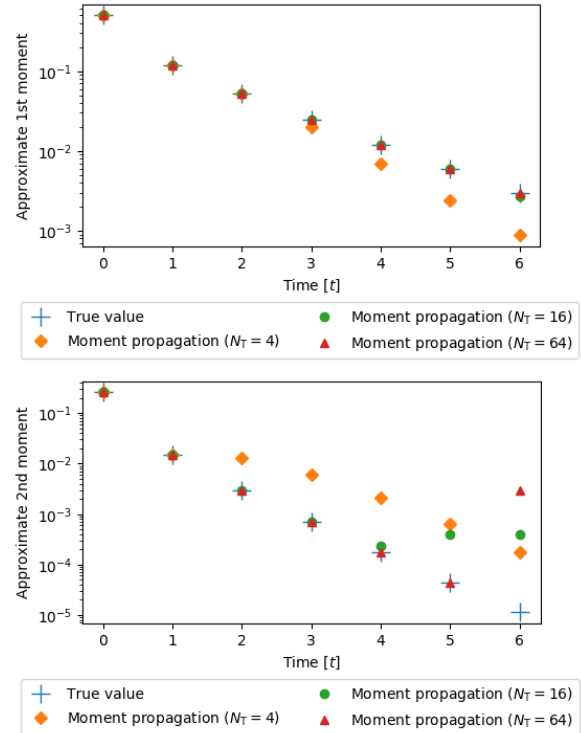


Fig. 2. Moment approximations for stochastic logistic map.

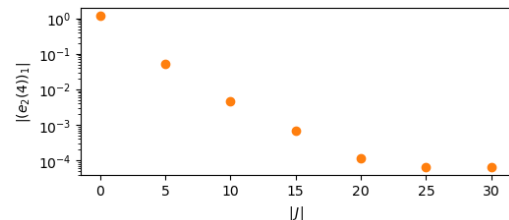


Fig. 3. Error bound on moment approximations.

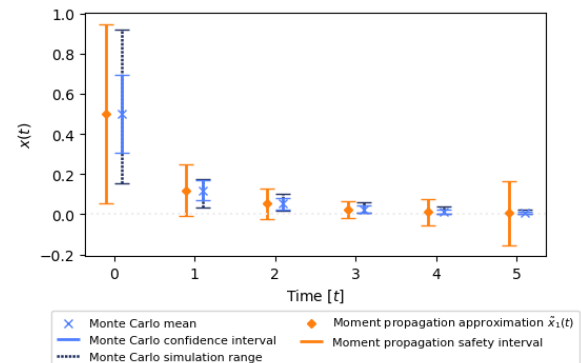


Fig. 4. Tail probability analysis.

predict their future positions. One way to achieve this is set-based reachability, as advocated by Althoff and Dolan (2014). To use their method, they must consider systems with bounded disturbances and use linearization around

an equilibrium, that is, they approximate a polynomial system by a linear one, using Lagrange remainders. Carleman linearization allows taking the effect of higher dimensions of the system into account more precisely than Lagrange remainders. Moreover, our approach is probabilistic, while theirs is set-based, so the two approaches give different types of guarantees.

We consider a scenario in which, at each time step, the autonomous vehicle measures its current position with some known sensor error distributions, computes moments of its current state based on this measurement and the known error distribution, and predicts its future positions (and moments) up to  $t$  steps ahead in time by applying the truncated system to these moments. This process is repeated forever, leading to the framework in Figure 5.

More precisely, we consider the *kinematic bicycle model* of a vehicle (Kong et al., 2015), which we rewrite as the following equivalent polynomial system

$$\begin{aligned} \dot{p}_x(t) &= v(t)c(t), & \dot{p}_y(t) &= v(t)s(t), \\ \dot{\psi}(t) &= \frac{v(t)}{\ell} \sin \beta, & \dot{v}(t) &= a(t), \\ \dot{c}(t) &= -\frac{s(t)v(t) \sin \beta}{\ell}, & \dot{s}(t) &= \frac{c(t)v(t) \sin \beta}{\ell}, \end{aligned}$$

for  $t \geq 0$ , where  $p_x(t) \in \mathbb{R}$  and  $p_y(t) \in \mathbb{R}$  represent the X–Y coordinates of the mass-center of the vehicle,  $v(t) \in \mathbb{R}$  denotes its speed,  $\psi(t)$  its inertial heading, and  $a(t) \in \mathbb{R}$  its acceleration. The constants  $\beta \in \mathbb{R}$  and  $\ell > 0$  respectively denote the angle of velocity and the distance from the vehicle’s rear axle to its mass-center. The scalars  $c(t)$  and  $s(t)$  are auxiliary variables that are introduced to obtain the polynomial model above from the original model of Kong et al. (2015) (which involves trigonometric terms), using the same techniques as Carothers et al. (2005).

The second-order Taylor expansion of the model above gives the following discrete-time approximation:

$$\begin{aligned} p_x(t + \Delta) &= p_x(t) + \Delta c(t)v(t) \\ &\quad + \frac{\Delta^2}{2} \left( a(t)c(t) - \frac{s(t)v^2(t) \sin \beta}{\ell} \right), \\ p_y(t + \Delta) &= p_y(t) + \Delta s(t)v(t) \\ &\quad + \frac{\Delta^2}{2} \left( a(t)s(t) + \frac{c(t)v^2(t) \sin \beta}{\ell} \right), \\ \psi(t + \Delta) &= \psi(t) + \Delta \frac{v(t)}{\ell} \sin \beta + \frac{\Delta^2}{2} \frac{a(t)}{\ell} \sin \beta, \\ v(t + \Delta) &= v(t) + \Delta a(t), \\ c(t + \Delta) &= c(t) - \Delta \frac{s(t)v(t) \sin \beta}{\ell} \\ &\quad - \frac{\Delta^2}{2} \left( \frac{c(t)v^2(t) \sin^2 \beta}{\ell^2} + \frac{a(t)s(t) \sin \beta}{\ell} \right), \\ s(t + \Delta) &= s(t) + \Delta \frac{c(t)v(t) \sin \beta}{\ell} \\ &\quad + \frac{\Delta^2}{2} \left( -\frac{s(t)v^2(t) \sin^2 \beta}{\ell^2} + \frac{a(t)c(t) \sin \beta}{\ell} \right), \end{aligned}$$

where  $\Delta > 0$ . To describe the evolution of the states of the vehicle at times  $0, \Delta, 2\Delta, \dots$ , we write this system in the form of (1). In particular, consider the discrete-time instant  $t \in \mathbb{N}$  corresponding to the continuous time  $t\Delta$ . By

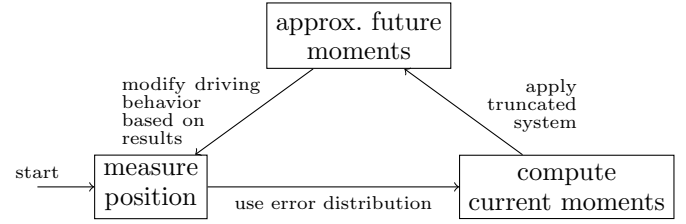


Fig. 5. Moment propagation framework for autonomous vehicles.

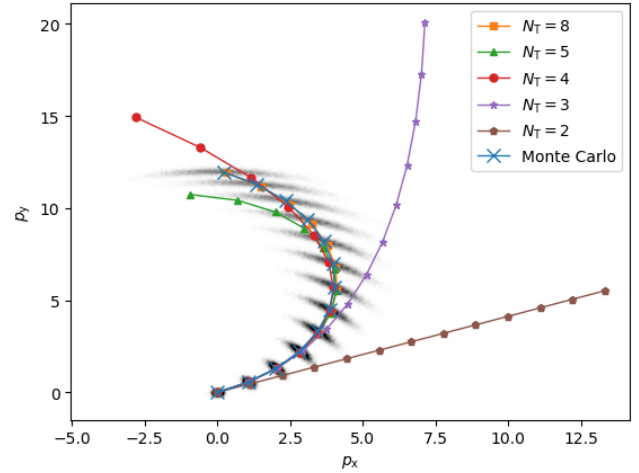


Fig. 6. First moment approximation of vehicle dynamics.

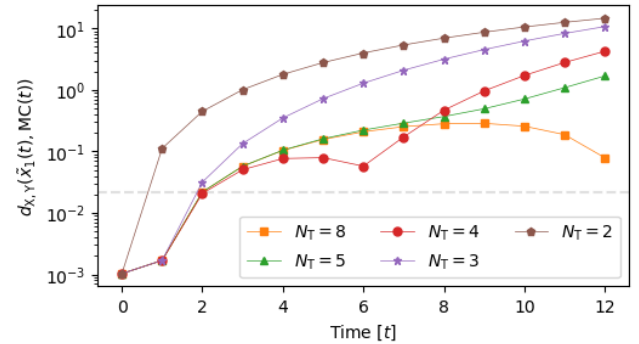


Fig. 7. Distance to the mean of the empirical distribution.

letting

$$x(t) \triangleq [p_x(t), p_y(t), \psi(t), v(t), c(t), s(t)]^T,$$

we obtain (1) where  $d_S = 3$  and the coefficients  $F_0(t), \dots, F_3(t)$  depend on  $\Delta, \beta, \ell$ , and  $a(t)$ . We consider the setting where the acceleration values  $a(0), a(1), \dots$  are independent uniform random variables over  $[0.9, 1]$ ,  $\Delta = 0.1$ ,  $\beta = \pi/8$ ,  $\ell = 2.5$ , and, for the initial state,  $p_x(0), p_y(0), v(0), \psi(0)$  are independent Gaussian random variables with mean 0 and standard deviation 0.1, and  $c(0) = \cos(\psi(0) + \beta)$  and  $s(0) = \sin(\psi(0) + \beta)$ .

**Experimental Results.** Figure 6 shows the expected trajectory of the vehicle as approximated by our method for different truncation limits, as well as the empirical distribution computed by 10000 runs of Monte Carlo simulation and the mean of that distribution. Figure 7 shows the distance  $d(\tilde{x}_1(t), MC(t))$  between  $\tilde{x}_1(t)$  and the mean of the empirical distribution for the same truncation limits.

Figures 6 and 7 show that larger truncation limits give systems that follow the empirical distribution better. It also shows that, for a fixed truncation limit, the distance to the empirical system grows larger with time. The only exception is around  $t = 6$  for  $N_T = 2$ , which can be explained by the fact that the trajectory of the truncated system crosses that of the empirical one. Notice that,  $N_T = 8$  gives an exact computation at  $t = 2$ , which means that the error at this point (corresponding to the horizontal line in Figure 7) is entirely induced by Monte Carlo. Therefore, it is pointless to consider errors that are of the same order as the one introduced by Monte Carlo at  $t = 2$ .

Also note that the largest truncation limit we use in this experiment is  $N_T = 8$ , which makes the online computation very fast.

## 5. CONCLUSION

We have explored the problem of computing moments of discrete-time polynomial stochastic systems. Through a truncated Carleman linearization approach, we have obtained an iterative procedure for approximating the moments of the state of the system across different time instants. Furthermore, we have provided tail-probability bounds to assess the safety properties of the system.

In our analysis, we have used spherical safety regions characterized using the Euclidean norm. By changing the norm to weighted norms induced by positive-definite matrices, ellipsoidal safety regions can also be considered. We note that safety analysis for a non-ellipsoidal region can be conducted by using ellipsoids that are inside that region. Finding the largest such ellipsoid allows us to do safety analysis in the least conservative way. This is possible by solving a matrix optimization problem, which we leave for future work.

One of our future research goals is to use the approximate moments to reconstruct an approximation of the distribution of the state of the system. There exist different techniques to achieve this, as given for example by Schmüdgen (2017) or John et al. (2007).

Another direction for future work is to make the algorithm more efficient by using the symmetry of Kronecker powers to reduce the sizes of the matrices involved in the computations, which would allow us to compute even higher moments and more time steps in the future.

## REFERENCES

- Al-Tuwaim, M.S., Crisalle, O.D., and Svoronos, S.A. (1998). Discretization of nonlinear models using a modified Carleman linearization technique. In *Proc. Amer. Contr. Conf.*, 3084–3088.
- Althoff, M. and Dolan, J.M. (2014). Online verification of automated road vehicles using reachability analysis. *IEEE Trans. Robotics*, 30(4), 903–918.
- Amini, A., Sun, Q., and Motee, N. (2019). Carleman state feedback control design of a class of nonlinear control systems. In *Proc. IFAC NecSys*.
- Athreya, K.B. and Dai, J. (2000). Random logistic maps. I. *J. Theor. Probab.*, 13(2), 595–608.
- Banzhaf, H., Dolgov, M., Stellet, J.E., and Zöllner, J.M. (2018). From footprints to beliefprints: Motion planning under uncertainty for maneuvering automated vehicles in dense scenarios. In *Proc. Int. Conf. Intel. Transport. Systems*, 1680–1687.
- Bellman, R. and Richardson, J.M. (1963). On some questions arising in the approximate solution of nonlinear differential equations. *Quart. Appl. Math.*, 20(4), 333–339.
- Bry, A. and Roy, N. (2011). Rapidly-exploring random belief trees for motion planning under uncertainty. In *Proc. IEEE ICRA*, 723–730.
- Cacace, F., Cusimano, V., Germani, A., and Palumbo, P. (2014). A Carleman discretization approach to filter nonlinear stochastic systems with sampled measurements. In *Proc. IFAC World Congr.*, 9534–9539.
- Cacace, F., Cusimano, V., Germani, A., Palumbo, P., and Papi, M. (2017). Optimal linear filter for a class of nonlinear stochastic differential systems with discrete measurements. In *Proc. IEEE Conf. Dec. Control*, 2807–2812.
- Carothers, D.C., Parker, G.E., Sochacki, J.S., and Warne, P.G. (2005). Some properties of solutions to polynomial systems of differential equations. *Electr. J. Differ. Eq.*, 2005(40), 1–17.
- Forets, M. and Pouly, A. (2017). Explicit error bounds for Carleman linearization. *Online*, <https://arxiv.org/abs/1711.02552>.
- Goswami, D. and Paley, D.A. (2017). Global bilinearization and controllability of control-affine nonlinear systems: A Koopman spectral approach. In *Proc. IEEE Conf. Dec. Contr.*, 6107–6112.
- Gray, H.L. and Wang, S. (1991). A general method for approximating tail probabilities. *J. Am. Stat. Assoc.*, 86(413), 159–166.
- Hashemian, N. and Armaou, A. (2019). Feedback control design using model predictive control formulation and Carleman approximation method. *AICHE J.*, 65, 1–11.
- John, V., Angelov, I., Öncül, A., and Thévenin, D. (2007). Techniques for the reconstruction of a distribution from a finite number of its moments. *Chem. Eng. Sci.*, 62(11), 2890–2904.
- Kong, J., Pfeiffer, M., Schildbach, G., and Borrelli, F. (2015). Kinematic and dynamic vehicle models for autonomous driving control design. In *Proc. IEEE Intell. Veh. Symp.*, 1094–1099.
- Laub, A.J. (2005). *Matrix Analysis for Scientists and Engineers*. SIAM.
- Mesbahi, A., Bu, J., and Mesbahi, M. (2019). On modal properties of the Koopman operator for nonlinear systems with symmetry. In *Proc. Amer. Contr. Conf.*, 1918–1923.
- Rauh, A., Minisini, J., and Aschemann, H. (2009). Carleman linearization for control and for state and disturbance estimation of nonlinear dynamical processes. In *Proc. IFAC Conf. MMR*, 455–460.
- Schmüdgen, K. (2017). *The Moment Problem*. Springer.
- Steeb, W.H. and Wilhelm, F. (1980). Non-linear autonomous systems of differential equations and Carleman linearization procedure. *J. Math. Anal. Appl.*, 77(2), 601–611.
- Wong, W.S. (1983). Carleman linearization and moment equations of nonlinear stochastic equations. *Stochastics*, 9(1-2), 77–101.