Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Data-driven Solution of Stochastic Differential Equations Using Maximum Entropy Basis Functions ⋆

Vedang M. Deshpande & Raktim Bhattacharya

Department of Aerospace Engineering, Texas A&M University
College Station, TX 77843, USA
(e-mail: vedang.deshpande@tamu.edu, raktim@tamu.edu)

Abstract: In this paper we present a data-driven approach for uncertainty propagation. In particular, we consider stochastic differential equations with parametric uncertainty. Solution of the differential equation is approximated using maximum entropy (maxent) basis functions similar to polynomial chaos expansions. Maxent basis functions are derived from available data by maximization of information-theoretic entropy, therefore, there is no need to specify basis functions beforehand. We compare the proposed maxent based approach with existing methods.

Keywords: uncertainty propagation, chaos expansion, maximum entropy, stochastic differential equations, data-driven models

1. INTRODUCTION

Most physical systems are modeled by non-linear differential equations. Such models often suffer from parametric uncertainties, which in turn leads to inaccurate prediction of the states of the system. Even if the state of a system is exactly known at a time instant, due to the parametric uncertainties, the states at subsequent time instants determined from uncertain models exhibit stochastic nature, and often one is interested in temporal evolution of distribution or statistical moments of the states.

In general, determining distribution of states, by propagation through stochastic models, is an infinite dimensional and computationally intractable problem. Polynomial chaos expansions (PCE) have been widely studied as an approach to construct a reduced order and computationally tractable surrogate model for the original system. In PCE, a stochastic process is expressed as a weighted sum of polynomials of random variables. The polynomials or basis functions are selected according to the underlying distribution of the random parameters, and the optimal weights or coefficients of the chaos are determined using the so called Galerkin projection. A correspondence between stochastic distributions and optimal basis from Askey family of orthogonal polynomials which leads to the exponential convergence of error with respect to the order of approximation is discussed in Xiu and Karniadakis (2002). This is also known as generalized polynomial chaos (gPC). Application of PCE for modeling uncertainties in physical systems has been discussed in many works, such as Xiu and Karniadakis (2003); Najm (2009); Hamdia et al. (2017), to name a few.

The classical polynomial chaos approach assumes that the distributions of involved random variables are exactly known. However, in reality, this assumption does not hold true. Often, limited information is available about the distributions in the form of finite number of samples, and engineering models must be developed using such limited set of data. A data-driven approach for chaos expansion, termed as arbitrary polynomial chaos (aPC), is discussed in Oladyshkin and Nowak (2012). In aPC, instead of fitting a probability distribution function over the given data, an orthogonal basis of polynomials for chaos expansion is constructed from raw moments of the available samples. In the present work, we pursue a similar approach, but instead of constructing polynomials from raw moments, we use maximum entropy (*maxent*) functions which are derived from the available samples, as basis to construct a chaos expansion.

The principle of maximum entropy was first introduced in Jaynes (1957a,b) as an approach to draw the least biased inference from incomplete or limited data. Maximum entropy based methods are often used for inferring a probability distribution from sparse information and it has found applications in many diverse fields such as structural mechanics, image processing, and machine learning, among others. See Duan et al. (2000); Karl (2005); Ziebart et al. (2008). The *maxent* basis functions that we use in this work arise as shape functions in the context of *maxent* based polygonal interpolation, as discussed in Sukumar (2004); Arroyo and Ortiz (2006). For the sake of completeness, *maxent* based polygonal interpolation is briefly reviewed in Section 2. Our recent work, Deshpande and Bhattacharya (2019b), discussed the data-driven function approximation using such *maxent* basis functions in the context of dynamics modeling. The results show that *maxent* based approach can model the unknown dynamics with good accuracy even if the available data set is sparse. This served as a motivation to pursue the present study.

⋆ This work was supported by the National Science Foundation (grant no. 1762825).

The objective of this work is to investigate the accuracy and convergence properties of chaos expansions constructed using *maxent* basis functions for solution of stochastic differential equations (SDEs).

The key highlights of this work are as follows.

- We present a novel approach for developing surrogate models of SDEs using data-driven *maxent* basis functions.
- The infinite dimensional solutions of SDEs are approximated in *maxent* based deterministic finite dimensional framework.
- Error and convergence properties of *maxent* based chaos expansions are studied by means of numerical simulations, and compared with the existing methods.
- Numerical results show that if the functional dependence of a system on the random variable is unknown, *maxent* based approach performs better than the polynomial expansions even if the available data is sparse.

The rest of the paper is organized as follows. Section 2 briefly discusses the principle of maximum entropy and the derivation of *maxent* basis functions. Solutions of SDEs using chaos expansions is reviewed in Section 3. Numerical results are presented in Section 4 followed by concluding remarks in Section 5.

2. MAXIMUM ENTROPY BASIS FUNCTIONS

2.1 Maximum entropy principle

Suppose Δ_i are mutually independent events with the unknown discrete probabilities p_i, such that $p_i \geq 0$ and p_i form the partition of unity, for $i = 1 \cdots N$. Let us assume that we have been given expected value of a function $\mathbb{E}[\mathbf{g}(\Delta)] := \sum_{i=1}^{N} p_i \mathbf{g}(\Delta_i)$. Our goal is to determine the probability distribution $\mathbf{p} := [p_1, p_2 \cdots p_N]^T$, which satisfies the given constraint. There can be a number of distinct distributions which satisfy the given constraints.

The maximum entropy principle serves as a tool to infer the least biased distribution from the insufficient data. To be specific, the *maxent* principle states that (see Jaynes (1957a,b)), out of all possible distributions that satisfy the given constraints, the least biased or the most likely probability distribution ($\mathbf{p}^*$), is the one which maximizes the information-theoretic entropy $H(\mathbf{p})$ defined as

$$H(\mathbf{p}) := -\sum_{i=1}^{N} p_i \log(p_i), \tag{1}$$

and $0 \log(0) := 0$. The entropy maximization problem, formally can be written as

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} H(\mathbf{p}), \tag{2a}$$

$$\text{such that } \sum_{i=1}^{N} p_i \mathbf{g}(\Delta_i) = \mathbb{E}[\mathbf{g}(\Delta)], \text{ and } \sum_{i=1}^{N} p_i = 1. \tag{2b}$$

Equivalently, the constraints in (2b) can be rewritten as

$$\sum_{i=1}^{N} p_i \big(\mathbf{g}(\Delta_i) - \mathbb{E}[\mathbf{g}(\Delta)]\big) = 0. \tag{3}$$

2.2 Minimum relative entropy principle

Suppose prior distribution $\mathbf{m} := [m_1, m_2 \cdots m_N]^T$ which estimates the probability distribution $\mathbf{p}$ is known, then the least biased probability distribution, $\mathbf{p}^*$, is determined by minimizing relative (cross) entropy or Kullback-Leibler divergence as (see Kullback and Leibler (1951))

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} \left(\bar{H}(\mathbf{p}, \mathbf{m}) := \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{m_i}\right) \right), \tag{4}$$

subject to constraints in (3).

As mentioned earlier, *maxent* basis functions that we use in the current work arise as shape functions in the context of polygonal interpolation, which is discussed next.

2.3 Polygonal interpolation using maxent principle

An analogy between determining the least biased probability distribution and polygonal interpolation was first presented in Sukumar (2004), followed by extensions of the work in Sukumar (2007); Arroyo and Ortiz (2006). We briefly discuss the polygonal interpolation using *maxent* basis functions below.

Let $\mathcal{S} := \{\boldsymbol{\Delta}_i\}_{i=1}^N \subset \mathcal{D}_{\boldsymbol{\Delta}} \subset \mathbb{R}^d$, be a set with cardinality N in $d-$dimensional space. Suppose that $\mathcal{D}_{\boldsymbol{\Delta}}$ is compact. Each element $\boldsymbol{\Delta}_i$ in $\mathcal{S}$ is associated with a *shape function* $\psi_i(\cdot) \geq 0$. The polygonal interpolant $\hat{f}(\cdot)$ of $f(\cdot)$ at $\boldsymbol{\Delta}$ is defined as

$$\hat{f}(\boldsymbol{\Delta}) = \sum_{i=1}^{N} \psi_i(\boldsymbol{\Delta}) f(\boldsymbol{\Delta}_i), \tag{5}$$

where $f(\cdot)$ is a scalar real valued function defined over $\mathcal{D}_{\boldsymbol{\Delta}}$, and $\boldsymbol{\Delta} \in \text{Conv}(\mathcal{S})$, and the convex hull of the set $\mathcal{S}$ is defined as

$$\text{Conv}(\mathcal{S}) := \{\boldsymbol{\Delta} | \boldsymbol{\Delta} = \sum_{i=1}^{N} w_i \boldsymbol{\Delta}_i, \sum_{i=1}^{N} w_i = 1, w_i \geq 0, \boldsymbol{\Delta}_i \in \mathcal{S}\}.$$

The values of $\boldsymbol{\Psi}(\boldsymbol{\Delta}) := [\psi_1(\boldsymbol{\Delta}), \psi_2(\boldsymbol{\Delta}) \cdots \psi_N(\boldsymbol{\Delta})]^T$ used in evaluation of the interpolant (5) are also referred to as barycentric coordinates of $\boldsymbol{\Delta}$ w.r.t. elements nodes $\boldsymbol{\Delta}_i$ of $\mathcal{S}$. In this paper, we refer elements $\boldsymbol{\Delta}_i \in \mathcal{S}$ as basis nodes, and associated shape functions $\psi_i(\cdot)$ as basis functions.

It is desirable for interpolants to recover constant and linear functions exactly. Therefore, following constraints are imposed on barycentric coordinates $\psi_i(\boldsymbol{\Delta})$ to guarantee the linear precision of the interpolant.

$$\sum_{i=1}^{N} \psi_i(\boldsymbol{\Delta}) = 1, \tag{6a}$$

$$\sum_{i=1}^{N} \psi_i(\boldsymbol{\Delta}) \boldsymbol{\Delta}_i = \boldsymbol{\Delta}, \text{ or, } \mathbf{S}\boldsymbol{\Psi}(\boldsymbol{\Delta}) = \boldsymbol{\Delta}, \tag{6b}$$

where $\mathbf{S} := [\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2 \cdots \boldsymbol{\Delta}_N]$. Or equivalently,

$$\tilde{\mathbf{S}}\boldsymbol{\Psi}(\boldsymbol{\Delta}) = \mathbf{0}, \tag{7}$$

where $\tilde{\boldsymbol{\Delta}}_i := \boldsymbol{\Delta}_i - \boldsymbol{\Delta}$ and $\tilde{\mathbf{S}} := [\tilde{\boldsymbol{\Delta}}_1, \tilde{\boldsymbol{\Delta}}_2 \cdots \tilde{\boldsymbol{\Delta}}_N]$.

After a direct comparison between (3) and (7), $\psi_i(\boldsymbol{\Delta})$ can be interpreted as probability associated with the *event*

$\mathbf{\Delta}_i$, and here $\mathbf{g}(\mathbf{\Delta}) = \mathbf{\Delta}$. Let $\mathbf{m}(\mathbf{\Delta}) := [m_1, m_2 \cdots m_N]^T$ be a suitably defined prior estimate for the barycentric coordinates $\mathbf{\Psi}(\mathbf{\Delta})$. Therefore, $\mathbf{\Psi}(\mathbf{\Delta})$ which minimizes the relative entropy defined by (4) can be interpreted as the least biased barycentric coordinates of $\mathbf{\Delta}$ w.r.t. basis nodes $\mathbf{\Delta}_i \in \mathcal{S}$ for a given prior estimate $\mathbf{m}(\mathbf{\Delta})$.

The relative entropy minimization problem is formally written as

$$\mathbf{\Psi}^*(\mathbf{\Delta}) = \underset{\mathbf{\Psi}(\mathbf{\Delta})}{\arg\min}\, \bar{H}\big(\mathbf{\Psi}(\mathbf{\Delta}), \mathbf{m}(\mathbf{\Delta})\big); \text{s.t. } \tilde{\mathbf{S}}\mathbf{\Psi}(\mathbf{\Delta}) = \mathbf{0} \quad (8)$$

Solution of this convex optimization problem using the method of Lagrange multipliers is discussed in Jaynes (1957a); Sukumar (2004); Arroyo and Ortiz (2006). This optimization problem reduces to the following system of nonlinear equations in Lagrange multipliers $\boldsymbol{\lambda}^T := [\lambda_1, \lambda_2 \cdots \lambda_d]$ associated with equality constraints in (8)

$$\sum_{i=1}^{N} \tilde{\mathbf{\Delta}}_i m_i(\mathbf{\Delta}) e^{-\boldsymbol{\lambda}^T \tilde{\mathbf{\Delta}}_i} = \mathbf{0}. \quad (9)$$

Numerical solvers are employed to solve this system of equations. The optimal barycentric coordinates or basis functions in terms of $\boldsymbol{\lambda}$ are given by

$$\psi_i^*(\mathbf{\Delta}) = \frac{m_i(\mathbf{\Delta}) e^{-\boldsymbol{\lambda}^T \tilde{\mathbf{\Delta}}_i}}{\sum_{j=1}^{N} m_j(\mathbf{\Delta}) e^{-\boldsymbol{\lambda}^T \tilde{\mathbf{\Delta}}_j}}. \quad (10)$$

In this paper, we use Gaussian prior, i.e. $\mathbf{m}(\mathbf{\Delta})$ for any $\mathbf{\Delta} \in \text{Conv}(\mathcal{S})$ is defined as

$$\mathbf{m}(\mathbf{\Delta}) := [m_1, m_2 \cdots m_N]^T \text{ and } m_i(\mathbf{\Delta}) := \frac{e^{-\beta||\tilde{\mathbf{\Delta}}_i||_2^2}}{\sum_j e^{-\beta||\tilde{\mathbf{\Delta}}_j||_2^2}},$$

where $\beta \geq 0$. For $\beta > 0$, we get so called *local maxent* basis functions. The parameter β affects the degree of locality of basis functions. Higher value of β implies the larger decay of basis functions away from their associated nodes, and larger degree of locality of basis functions. For $\beta = 0$ or uniform prior, i.e. $m_i = 1/N$ for $i = 1 \cdots N$, we recover so called *global maxent* basis functions introduced in Sukumar (2004). All numerical results presented in this paper are obtained for $\beta = 0$. For notational convenience, hereafter we drop the asterisk in (10) and use $\mathbf{\Psi}(\mathbf{\Delta})$ or $\psi_i(\mathbf{\Delta})$ to denote the optimal local *maxent* basis functions evaluated at $\mathbf{\Delta}$.

3. SOLUTION OF STOCHASTIC DIFFERENTIAL EQUATIONS

Let $\mathbf{\Delta} \in \mathcal{D}_{\mathbf{\Delta}} \subset \mathbb{R}^d$ be a random vector with joint probability density function $p(\mathbf{\Delta})$. Approximant of a function $\mathbf{f}(\mathbf{\Delta}) : \mathcal{D}_{\mathbf{\Delta}} \mapsto \mathbb{R}^n$ using a known set of basis functions $\{\phi_i(\mathbf{\Delta})\}_{i=1}^N$ can be written as

$$\mathbf{f}(\mathbf{\Delta}) \approx \hat{\mathbf{f}}(\mathbf{\Delta}) := \sum_{i=0}^{N} \mathbf{f}_i \phi_i(\mathbf{\Delta}), \quad (11)$$

where, $\mathbf{f}_i$ are deterministic coefficients. The optimal coefficients are determined by Galerkin projection, i.e. the projection of the error $\mathbf{e}(\mathbf{\Delta}) := \mathbf{f}(\mathbf{\Delta}) - \hat{\mathbf{f}}(\mathbf{\Delta})$ against each basis function is set to zero, i.e.

$$\mathbb{E}[\mathbf{e}(\mathbf{\Delta})\phi_i(\mathbf{\Delta})] := \int_{\mathcal{D}_{\mathbf{\Delta}}} \mathbf{e}(\mathbf{\Delta})\phi_i(\mathbf{\Delta}) p(\mathbf{\Delta}) d\mathbf{\Delta} = \mathbf{0}, \quad (12)$$

for $i = 0, \cdots, N$. This results in a deterministic system of equations which can be solved for $\mathbf{f}_i$. If instead of $p(\mathbf{\Delta})$, samples of $\mathbf{\Delta}$ are available, then integral in (12) can be replaced with summation and it becomes a least squares problem.

In generalized polynomial chaos (gPC) expansions, $\phi_i(\cdot)$ are multivariate orthogonal polynomials which are selected based on the known distribution of $\mathbf{\Delta}$, see Xiu and Karniadakis (2002). However, if $p(\mathbf{\Delta})$ is unknown, instead, samples of random vector $\mathbf{\Delta}$ are available then one has to rely on data-driven methods. A data-driven polynomial chaos approach in which orthogonal basis polynomials are constructed from raw moments of available samples, termed as arbitrary polynomial chaos (aPC), is presented in Oladyshkin and Nowak (2012). In present work, we employ *maxent* functions derived in Section 2 as bases for chaos expansions. Data-driven function approximation using *maxent* basis functions is discussed in detail in Deshpande and Bhattacharya (2019b). In this paper, we focus on the application of *maxent* basis functions for solution of stochastic differential equations.

A general procedure for solution of stochastic differential equations using chaos expansions is discussed in Xiu and Karniadakis (2002). In this paper, for simplicity and brevity of discussion, and without loss of generality, we consider linear stochastic ordinary differential equation given by

$$\dot{\mathbf{x}}(t, \mathbf{\Delta}) = \mathbf{A}(\mathbf{\Delta})\mathbf{x}(t, \mathbf{\Delta}), \quad (13)$$

where $\mathbf{x} := \mathbf{x}(t, \mathbf{\Delta}) \in \mathbb{R}^n$ and $\mathbf{A}(\mathbf{\Delta}) \in \mathbb{R}^{n \times n}$. The solution $\mathbf{x}(t, \mathbf{\Delta})$ is approximated using chaos expansion as

$$\mathbf{x}(t, \mathbf{\Delta}) \approx \hat{\mathbf{x}}(t, \mathbf{\Delta}) := \sum_{i=0}^{N} \mathbf{x}_i(t)\phi_i(\mathbf{\Delta}) =: \mathbf{X}(t)\mathbf{\Phi}(\mathbf{\Delta}), \quad (14)$$

where $\mathbf{X}(t) := [\mathbf{x}_1(t) \cdots \mathbf{x}_N(t)]$ are time varying deterministic coefficients and $\mathbf{\Phi}(\mathbf{\Delta}) := [\phi_1(\mathbf{\Delta}) \cdots \phi_N(\mathbf{\Delta})]^T$ are chosen basis functions for chaos expansion. After substituting the approximation (14), the projection of equation error of (13) against each basis is set to zero to obtain a system of deterministic ODEs in terms of chaos coefficients $\mathbf{x}_i(t)$, which is given by

$$\dot{\mathbf{x}}_c = \big(\mathbb{E}[\mathbf{\Phi}(\mathbf{\Delta})\mathbf{\Phi}^T(\mathbf{\Delta})] \otimes \mathbf{I}_n\big)^{-1} \times \\ \mathbb{E}[(\mathbf{\Phi}(\mathbf{\Delta})\mathbf{\Phi}^T(\mathbf{\Delta})) \otimes \mathbf{A}(\mathbf{\Delta}))]\mathbf{x}_c, \quad (15)$$

where, $\mathbf{x}_c(t) := \mathbf{vec}\,(\mathbf{X}(t))$. A step-by-step derivation of the surrogate system of deterministic ODEs similar to (15) for a generalized non-linear ODE can be found in Deshpande and Bhattacharya (2019a).

The deterministic ODEs in (15) can be solved using standard deterministic approaches. Therefore, $\mathbf{x}_c(t)$ and hence $\mathbf{X}(t)$ can be determined at any time t. Then the moments of $\mathbf{x}(t, \mathbf{\Delta})$ can be approximated by the moments of $\hat{\mathbf{x}}(t, \mathbf{\Delta})$ as

$$\mathbb{E}[\mathbf{x}(t)] \approx \mathbb{E}[\hat{\mathbf{x}}(t)] = \mathbf{X}(t)\mathbb{E}[\mathbf{\Phi}(\mathbf{\Delta})], \quad (16a)$$

$$\mathbb{E}\big[\big(\mathbf{x}(t) - \mathbb{E}[\mathbf{x}(t)]\big)\big(\,\cdot\,\big)^T\big] \approx \mathbb{E}\big[\big(\hat{\mathbf{x}}(t) - \mathbb{E}[\hat{\mathbf{x}}(t)]\big)\big(\,\cdot\,\big)^T\big] \\ = \mathbf{X}(t)\Big(\mathbb{E}[\mathbf{\Phi}(\mathbf{\Delta})\mathbf{\Phi}^T(\mathbf{\Delta})] - \mathbb{E}[\mathbf{\Phi}(\mathbf{\Delta})]\mathbb{E}[\mathbf{\Phi}(\mathbf{\Delta})]^T\Big)\mathbf{X}^T(t). \quad (16b)$$

Let $\mathcal{D} := \{\mathbf{\Delta}_i\}_{i=1}^{n_\mathcal{D}} \subset \mathcal{D}_{\mathbf{\Delta}}$ be the set of available samples for $\mathbf{\Delta}$. Since this is the only available information about

$\mathbf{\Delta}$, the expectation integrals involved in (15) and (16) are approximated by sample averages calculated using data points in $\mathcal{D}$.

We define $\mathcal{B} := \{\mathbf{\Delta}_i\}_{i=1}^{n_\mathcal{B}}$ such that $\mathcal{D} \subseteq \text{Conv}(\mathcal{B})$. This is trivially satisfied if we choose $\mathcal{B} = \mathcal{D}$. However, $\mathcal{B}$ need not be identical to $\mathcal{D}$. The elements in $\mathcal{B}$ serve as nodes for *maxent* basis functions. Therefore, the number of *maxent* basis functions is $n_\mathcal{B}$. The definition of $\mathcal{B}$ ensures that barycentric coordinates (see (10)) of each $\mathbf{\Delta}_i \in \mathcal{D}$ are well defined w.r.t. basis nodes in $\mathcal{B}$.

In the following section we solve different stochastic ODEs using *maxent* basis functions as bases for chaos expansions.

4. NUMERICAL RESULTS

For the purpose of numerical simulations, we consider a scalar ODE given by

$$\dot{x}(t) = a(\Delta)x(t), \quad \Delta \in \mathbb{R}, \tag{17}$$

with deterministic initial condition $x(0) = 1$. The decay coefficient $a(\Delta)$ is a stochastic parameter with dependence on random variable Δ. The system of ODEs for chaos coefficients when using *maxent* basis functions follows from (14) and (15) as

$$x(t,\Delta) \approx \hat{x}(t,\Delta) = \sum_{i=0}^{n_\mathcal{B}} x_i(t)\psi_i(\Delta),$$

$$\dot{\mathbf{x}}_c = \left(\mathbb{E}[\mathbf{\Psi}(\Delta)\mathbf{\Psi}^T(\Delta)]\right)^{-1}\mathbb{E}[a(\Delta)\mathbf{\Psi}(\Delta)\mathbf{\Psi}^T(\Delta)]\mathbf{x}_c, \tag{18}$$

where $\mathbf{x}_c(t) = [x_1(t), x_2(t)\cdots x_{n_\mathcal{B}}(t)]^T$. Note, $\mathbf{\Psi}(\Delta)$ are barycentric coordinate of Δ w.r.t basis nodes in $\mathcal{B}$. Approximate mean and variance at time t follow from (16) as

$$\hat{\mu}_x(t) = \mathbf{x}_c(t)^T\mathbb{E}[\mathbf{\Psi}(\Delta)], \tag{19a}$$

$$\hat{\sigma}_x^2(t) = \mathbf{x}_c^T(t)(\mathbb{E}[\mathbf{\Psi}(\Delta)\mathbf{\Psi}^T(\Delta)] - \mathbb{E}[\mathbf{\Psi}(\Delta)]\mathbb{E}[\mathbf{\Psi}(\Delta)]^T)\mathbf{x}_c(t) \tag{19b}$$

Since *maxent* basis functions evaluated for any Δ add up to one (see (6a)), the initial condition for chaos coefficients can be written as $\mathbf{x}_c(0) = \mathbf{1}_{n_\mathcal{B}}$, where, $\mathbf{1}_{n_\mathcal{B}}$ is the column vector of length $n_\mathcal{B}$ with each element unity. With this initial condition for $\mathbf{x}_c$, it is straightforward to verify that mean and variance approximated using (19) come out to be $\hat{\mu}_x(0) = 1$ and $\hat{\sigma}_x^2(0) = 0$, which exactly recovers the given initial condition $x(0) = 1$.

In text to follow, we consider two examples. The first example is a classical ODE studied in chaos expansion literature. In this example, we assume that the functional form of $a(\Delta)$ is known and investigate the error convergence properties of *maxent* based expansions. In second example, we assume that the functional form of $a(\Delta)$ is unknown, instead, an approximant is constructed from a given sparse data set. Then a surrogate model is developed using the approximant. The results obtained for *maxent* based models are compared with data-driven polynomial chaos (aPC) based surrogate models.

Example 1

Let us assume that the functional form of $a(\Delta) = -(1 + \Delta)/2$ is known, where Δ is uniformly distributed over $[-1, 1]$. This ODE has been used as a test problem in Xiu and Karniadakis (2002); Oladyshkin and Nowak (2012).

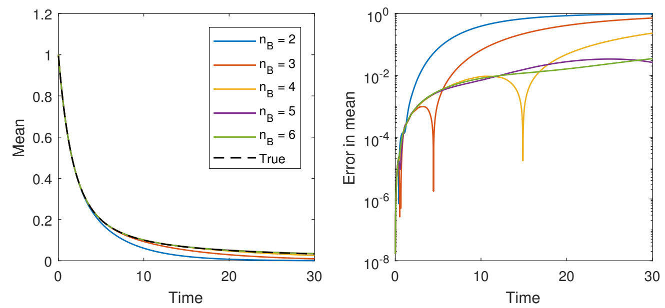

Fig. 1. Estimated mean and error for different number of *maxent* basis functions ($n_\mathcal{B}$) and fixed $n_\mathcal{D} = 500$.

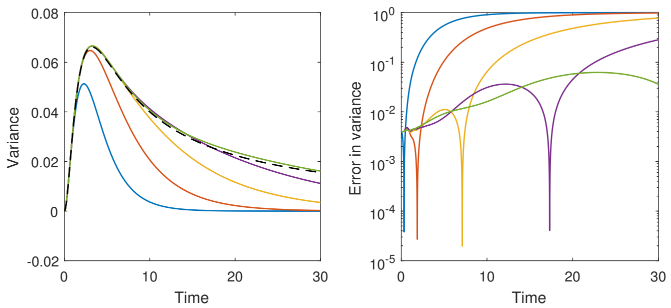

Fig. 2. Estimated variance and error for different number of *maxent* basis functions ($n_\mathcal{B}$) and fixed $n_\mathcal{D} = 500$.

The analytical or true values of mean and variance are given by $\mu_x(t) = \frac{1-e^{-t}}{t}$ and $\sigma_x^2(t) = \frac{1-e^{-2t}}{2t} - \left(\frac{1-e^{-t}}{t}\right)^2$ for $t > 0$, which are shown by black dashed lines in Fig.1 and Fig.2.

The two major sources of error in estimated moments given by (19) are namely, limited number of data samples ($n_\mathcal{D}$), and truncation of chaos expansion at finite number of terms or finite number of basis functions ($n_\mathcal{B}$). Convergence properties of both types of error are discussed in the following text. We first consider the effect the of number of basis functions, and for this investigation we assume the data set $\mathcal{D}$ to be fixed. For the second case, we keep the basis functions fixed, and vary the size of data set $\mathcal{D}$.

Fig.1 and Fig.2 show the effect of increasing number of *maxent* basis functions for a fixed data set $\mathcal{D}$ which consists of $n_\mathcal{D} = 500$ uniformly spaced points in $[-1, 1]$. The set of basis nodes, $\mathcal{B}$, is selected as a collection of $n_\mathcal{B}$ uniformly spaced points in $[-1, 1]$. Equation (18) is integrated numerically from $t = 0$ to $t = 30$ for different values of $n_\mathcal{B}$. Moments at each time instant are approximated using (19) and are shown in plots on the left sides of Fig.1 and Fig.2. It is well known that moments approximated using surrogate model such as (18) deviate from true values as ODEs are integrated forward in time. The errors in estimated moments calculated w.r.t. true moments are shown in plots on the right sides of Fig.1 and Fig.2. The error in estimated mean is defined as $\varepsilon_\mu(t) := |1 - \hat{\mu}_x(t)/\mu_x(t)|$. The error in estimated variance is defined in similar way. It is clear from Fig.1 and Fig.2 that surrogate models obtained using more number of *maxent* basis functions maintain better accuracy for longer periods of time while performing the temporal integration.

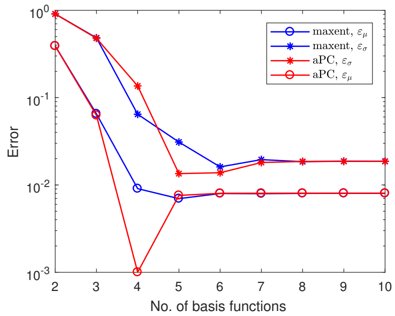

Fig. 3. Error in estimated moments at $t = 10$ for different number of basis functions ($n_\mathcal{B}$) and fixed $n_\mathcal{D} = 500$.

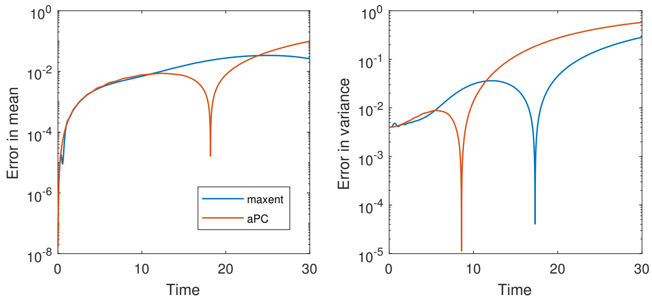

Fig. 4. Comparison of error in estimated moments for aPC and *maxent* based models, $n_\mathcal{B} = 5$ and $n_\mathcal{D} = 500$.

Fig.3 shows the convergence of error in moments calculated at $t = 10$ w.r.t. number of basis functions ($n_\mathcal{B}$). It is clear that both *maxent* and aPC based models demonstrate similar error convergence trend. We observe strong convergence initially as we start increasing number of basis functions. However, after certain value of $n_\mathcal{B}$, the errors saturate and do not decreases further as $n_\mathcal{B}$ is increased. For example, consider error in mean shown by blue line with circles in Fig.3. We observe strong convergence as the number of basis functions is increased from $n_\mathcal{B} = 2$ to $n_\mathcal{B} = 4$. However, for $n_\mathcal{B} = 5$ onwards, the errors become stagnant. This observation is attributed to limited size of sample set. This suggests that given a finite set of samples, $\mathcal{D}$, there is a lower limit on how much error can be reduced.

Fig.4 compares temporal evolution of error in estimated moments for data-driven surrogate models obtained using aPC and the proposed *maxent* based framework, for equal number of basis functions. We observe that *maxent* and aPC based models demonstrate similar error evolution, with *maxent* based model being marginally better than the aPC based model for estimated variance for $t > 10$.

Needless to say, the accuracy of approximation depends on the number of available samples, i.e. $n_\mathcal{D}$. The left plot in Fig.5 shows the mean of errors in estimated mean and variance of $x(t)$ at $t = 10$ calculated for different number of data points ($n_\mathcal{D}$) sampled randomly from $[-1, 1]$, repeated 500 times, for a fixed number of *maxent* basis functions $n_\mathcal{B} = 5$. Similar variation for the variance of errors in estimated moments is shown in the right plot. With no surprise, we observe that both mean and variance of errors decrease as the number of available samples increases. The aPC based model also shows a very similar trend (not shown in the figure).

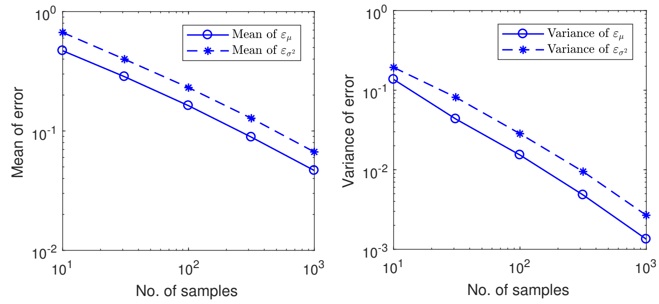

Fig. 5. Variation of mean (left) and variance (right) of error in estimated moments with different number of samples ($n_\mathcal{D}$), for fixed $n_\mathcal{B} = 5$ *maxent* basis functions, at $t = 10$.

Results discussed herein are obtained if Δ is sampled from a uniform distribution. Although not shown here, similar results are observed when Δ is sampled from a Gaussian distribution. In this example, where functional form of $a(\Delta)$ is assumed to be known, we observe that aPC and *maxent* based models demonstrate similar performances in terms of accuracy of the estimated moments and error convergence rates.

Example 2

In this example we assume that the functional form of $a(\Delta)$ is not known. Instead, we have been given samples of Δ, and values of $a(\Delta)$ for a sparse sample set of Δ. We assume that $n_\mathcal{D}$ samples of Δ are given. Let $\mathcal{D}' := \{\Delta_j\}_{j=1}^{n_{\mathcal{D}'}}$ be the sparse set of samples of Δ for which $a_j := a(\Delta_j)$ is known.

Here, solving stochastic ODE involves two steps. First, we approximate the function $a(\Delta)$ using given data $\{\Delta_j, a_j\}_{j=1}^{n_{\mathcal{D}'}}$. Second, we construct chaos expansion as we did in the previous example and integrate the surrogate model temporally. Basis nodes for *maxent* basis functions are chosen to be the elements in $\mathcal{D}'$, i.e. $\mathcal{B}$ and $\mathcal{D}'$ are identical and $n_\mathcal{B} = n_{\mathcal{D}'}$. The same basis nodes are used for function approximation in step one and chaos expansion in step two. The weights or coefficients associated with basis functions for function approximation are determined by the least squares solution as discussed in Deshpande and Bhattacharya (2019b).

For the purpose of simulation, we assume that Δ is a uniformly distributed variable between $(0, 1)$, and that $\mathcal{D}$ and $\mathcal{D}'$ respectively consist of uniformly spaced $n_\mathcal{D} = 500$ and $n_{\mathcal{D}'} = 10$ points in $(0, 1)$. The unknown underlying functional form of $a(\Delta)$ is assumed to be $a(\Delta) = \Delta^{\alpha-1}(1-\Delta)^{\gamma-1}$ with $\alpha = 2, \gamma = 0.5$. Moments of the true or reference solution of (17) are obtained using Monte-Carlo runs with 5×10^4 samples of Δ.

Comparison of error in estimated moments using *maxent* and aPC based models is shown in Fig.6, and clearly, the former demonstrates better accuracy. The relatively better accuracy of the *maxent* model is mainly attributed to the ability of *maxent* basis functions to approximate the underlying functional form of $a(\Delta)$ with sparse data better than polynomials bases constructed for the aPC model.

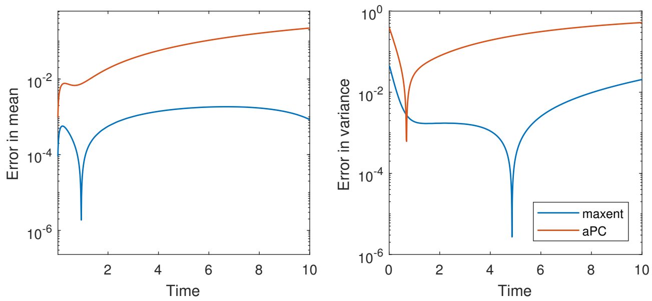

Fig. 6. Comparison of error in estimated moments, $n_{\mathcal{B}} = n_{\mathcal{D}'} = 10$, $n_{\mathcal{D}} = 500$.

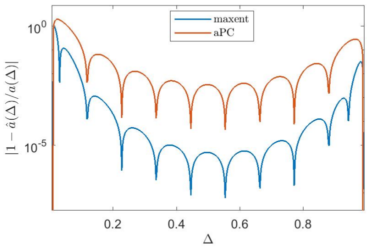

Fig. 7. Comparison of function approximation accuracy, $a(\Delta) = \Delta/\sqrt{1-\Delta}$, $n_{\mathcal{B}} = n_{\mathcal{D}'} = 10$.

The normalized error for function approximation is shown in Fig.7. The *maxent* based approximant has lower error than the aPC based function approximant, and consequently, *maxent* based model demonstrates better accuracy in the estimated moments as shown in Fig.6. However, as discussed in Deshpande and Bhattacharya (2019b), it should be noted that if the underlying functional form of $a(\Delta)$ lies within the span of basis polynomials, then aPC based model will have similar, or perhaps even better accuracy than the *maxent* model. But, for functions such as we considered here, which are difficult to approximate using polynomials, especially if the given data set is sparse, *maxent* based models perform better. Ability to accurately approximate functions is particularly vital for solving differential equations with the initial condition uncertainty, as modeling of uncertainty in the initial condition reduces to a function approximation problem.

5. CONCLUSION

In this paper we presented an approach to develop surrogate models based on limited data by construction of chaos expansions using *maxent* basis functions. We investigated the error characteristics and convergence properties of such *maxent* based chaos expansions and compared the accuracy with data-driven polynomial chaos expansions (aPC). We observed that *maxent* based surrogate models demonstrate similar accuracy as the aPC models if the functional dependence on the random variables is known. In the case where functional dependence is unknown, *maxent* based approach demonstrated better accuracy than the polynomial expansions especially if the available data is sparse.

We note that the selection of basis nodes for *maxent* basis functions may not be trivial for high dimensional complex systems, and it may affect the accuracy of chaos expansion. However, this investigation is out of scope of this paper. In this paper, we considered simple differential equations, with data set sampled from uniform and Gaussian distributions. However, results observed in this paper are motivating enough to pursue further investigations with more complex underlying distributions and differential equations of real physical systems, and will be topics of our future works.

REFERENCES

Arroyo, M. and Ortiz, M. (2006). Local maximum-entropy approximation schemes: a seamless bridge between finite elements and meshfree methods. *Int J Numer Methods Eng*, 65(13), 2167–2202.

Deshpande, V.M. and Bhattacharya, R. (2019a). On improved statistical accuracy of low-order polynomial chaos approximations. ArXiv:1909.03516.

Deshpande, V.M. and Bhattacharya, R. (2019b). Surrogate modeling of dynamics from sparse data using maximum entropy basis functions. ArXiv:1911.03016.

Duan, B., Templeman, A., and Chen, J. (2000). Entropy-based method for topological optimization of truss structures. *Computers & Structures*, 75(5), 539–550.

Hamdia, K.M., Silani, M., Zhuang, X., He, P., and Rabczuk, T. (2017). Stochastic analysis of the fracture toughness of polymeric nanoparticle composites using polynomial chaos expansions. *International Journal of Fracture*, 206(2), 215–227.

Jaynes, E.T. (1957a). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630.

Jaynes, E.T. (1957b). Information theory and statistical mechanics ii. *Physical Review*, 108(2), 171–190.

Karl, W.C. (2005). *Handbook of Image and Video Processing*, chapter 3.6 Regularization in Image Restoration and Reconstruction, 183–202. Elsevier.

Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Stat.*, 22(1), 79–86.

Najm, H.N. (2009). Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 41, 35–52.

Oladyshkin, S. and Nowak, W. (2012). Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliability Engineering and System Safety*, 106, 179–190.

Sukumar, N. (2004). Construction of polygonal interpolants: a maximum entropy approach. *Int J Numer Methods Eng*, 61(12), 2159–2181.

Sukumar, N. (2007). Overview and construction of meshfree basis functions: From moving least squares to entropy approximants. *Int J Numer Methods Eng*, 70(2), 181–205.

Xiu, D. and Karniadakis, G.E. (2002). The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2), 619–644.

Xiu, D. and Karniadakis, G.E. (2003). Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, 187(1), 137–167.

Ziebart, B.D., Maas, A., Bagnell, J.A., and Dey, A.K. (2008). Maximum entropy inverse reinforcement learning. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*.