Nonlinear Estimation in Polynomial Chaos Framework

Parikshit Dutta\textsuperscript{a}, Raktim Bhattacharya\textsuperscript{b}

\textsuperscript{a}Graduate Student, Texas A\&M University, College Station, TX
\textsuperscript{b}Assistant Professor, Texas A\&M University, College Station, TX

Abstract

In this paper we present two nonlinear estimation algorithms that combine generalized polynomial chaos theory with higher moment updates and Bayesian framework. Polynomial chaos theory is used to predict the evolution of uncertainty of the nonlinear random process. In the first estimation algorithm, higher order moment updates are used to estimate the posterior non Gaussian probability density function of the random process. The moments are updated using a linear gain. In the second method, Bayesian update rule is used to determine the posterior probability density function. Since polynomial chaos is a method of moments, it does not directly yield the needed probability density functions. The density function is determined from the moments using Gaussian mixture models and maximum entropy optimization. The nonlinear estimation algorithms are applied to the duffing oscillator system with initial condition uncertainty and its performance is compared with an estimator based on extended Kalman filtering (EKF) framework. We observe that the proposed estimators outperform the EKF based estimator when measurements are not available very frequently, thus highlighting the need for nonlinear estimator in such scenarios.

Key words: Nonlinear Estimation, Polynomial Chaos, Higher Order Moments, Bayesian Framework.

1 Introduction

The scientific community is currently facing challenges in understanding the impact of uncertainty in complex physical phenomenon and engineering systems. These problems arise from diverse applications including structural mechanics (e.g. excitation of a building due to a seismic event), petroleum engineering (e.g. effect of uncertainty in the porosity of the subterranean reservoir, on the flow of oil), astrodynamics (e.g. trajectory estimation of 99942 Apophis and space situational awareness in general), flight mechanics (e.g. reentry of space vehicles in uncertain atmosphere of Mars) and robotics (e.g. motion planning in uncertain environments). All these problems are characterized by uncertainty in the physical models and their parameters. Estimation of parameters for these systems is a hard problem because of the nonlinearities in the system and the lack of frequent measurements. Thus the evolution of uncertainty, which can be non Gaussian, needs to be predicted over longer intervals of time. These issues undermine the validity of the classical linear Gaussian theory. The main challenges of nonlinear estimation are in the prediction of uncertainty for nonlinear dynamical systems and its correction based on measurements.

Parameter estimation is usually performed in the Bayesian framework where uncertainty in the parameter is represented in terms probability density functions. Bayesian parameter estimation for linear Gaussian systems is optimal with Kalman filters \cite{1}. For nonlinear systems exhibiting Gaussian behaviour, the nonlinear system is linearized locally, about the current mean, and the covariance is propagated using the approximated linear dynamics. This method is used in extended Kalman filters (EKF) \cite{2}. It is well known that this approach performs poorly when nonlinearities are high resulting in unstable estimator \cite{3–6}. However, the error in mean and covariance can be reduced if the uncertainty is propagated, using the nonlinear dynamics, for a minimal set of sample points, called sigma points. The density function of the state is parameterized by the sigma points, which completely capture the true mean and

\textit{Email addresses: p0d5585@aero.tamu.edu (Parikshit Dutta), raktim@aero.tamu.edu (Raktim Bhattacharya).}
covariance. This approximated density function, when propagated through the true non-linear system, captures the posterior mean and covariance accurately to the third order (Taylor series expansion) for any nonlinearity with Gaussian behaviour. This technique has resulted in the unscented Kalman filter (UKF) [7]. Recently a simulation based estimation algorithm has been proposed [8], which are ensemble-based assimilation methods that are not limited to Gaussian evolution of the probability density function. Since these algorithms are simulation based, they do not scale well with dimension of the system, in terms of computational complexity [9]. In this paper, we use polynomial chaos (PC) theory to propagate uncertainty in nonlinear dynamical system, and develop two estimation algorithms using higher order moment updates and Bayesian framework. Since polynomial chaos predicts propagation of uncertainty in a more computationally efficient manner than Monte-Carlo [32], it is expected that nonlinear estimators based on polynomial chaos will be more efficient than particle filters. However, the computational complexity of polynomial chaos grows factorially with the parameter dimension. Thus, the proposed estimation algorithms are best suited for nonlinear systems with few parameters and slow measurement updates.

The PC framework can be thought of as an extension of Volterra’s theory of nonlinear functionals [10] for stochastic systems. Polynomial chaos was first introduced by Wiener [11] where Hermite polynomials were used to model stochastic processes with Gaussian random variables. According to Cameron and Martin [12] such an expansion converges in the $L_2$ sense for any arbitrary stochastic process with finite second moment. This applies to most physical systems. Xiu et al.[13] generalized the result of Cameron-Martin to various continuous and discrete distributions using orthogonal polynomials from the so called Askey-scheme [14] and demonstrated $L_2$ convergence in the corresponding Hilbert functional space. This is popularly known as the generalized polynomial chaos (gPC) framework. The gPC approach is limited to parametric uncertainty only and cannot be used to study the effect of stochastic forcing directly. However, for stochastic forcing processes that have nonzero correlation window, the process can be approximated to arbitrary accuracy using the Karhunen-Loeve expansions. This approximates the random process using known functions of time with coefficients that are random variables, thus transforming it to a problem with parametric uncertainty. The problem then becomes amenable for analysis in the gPC framework. The gPC framework has been applied to various applications including stochastic fluid dynamics [15,16], stochastic finite elements [17], and solid mechanics [18,19]. Application of gPC to problems related to control and estimation of dynamical systems, has been surprisingly limited [20–25].

In the context of nonlinear estimation, polynomial chaos has been applied by Blanchard et al.[26,27], where uncertainty prediction was computed using gPC theory for nonlinear dynamical systems, and estimation was performed using linear output equations and classical Kalman filtering theory. The work of Julier et al.[28] introduced estimation using higher order moment updates, while assuming that the estimated process is Gaussian at any given time. Majji et al.[29] extended the idea of estimation with higher order moments to non Gaussian processes, where the nonlinear dynamics was approximated using Taylor series approximations. The benefit of the gPC theory is that the computational complexity for uncertainty propagation and moment update is much lower than with Taylor series approximation of the dynamics and output equation. The latter approach results in ordinary differential equations involving tensors. As mentioned before, the limitation of gPC theory is that it can only incorporate parametric uncertainty in the system, which is the premise of the estimation algorithm presented here. The second estimation algorithm, based on Bayesian framework is more efficient than the estimators with higher order moment updates as tensor calculations require for moments updates are not needed here. The posterior probability density function is determined from the likelihood functions.

The paper is organized as follows. We first introduce polynomial chaos theory and demonstrate how stochastic differential equations can be solved in this framework. This is followed by the details of estimation algorithm that incorporate gPC theory and higher order moment updates. The paper concludes with an example that demonstrates performance on this estimator in comparison with the linear estimators developed in EKF framework.

## 2 Wiener-Askey Polynomial Chaos

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where $\Omega$ is the sample space, $\mathcal{F}$ is the $\sigma$-algebra of the subsets of $\Omega$, and $\mathcal{M}$ is the probability measure. Let $\Delta(\omega) = (\Delta_1(\omega), \cdots, \Delta_d(\omega)) : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}^d)$ be an $\mathbb{R}^d$-valued continuous random variable, where $d \in \mathbb{N}$, and $\mathcal{B}^d$ is the $\sigma$-algebra of Borel subsets of $\mathbb{R}^d$. A general second order process $X(\omega) \in L_2(\Omega, \mathcal{F}, \mathcal{M})$ can be expressed by polynomial chaos as

$$X(\omega) = \sum_{i=0}^{\infty} x_i \phi_i(\Delta(\omega)), \quad (1)$$

where $\phi_i(\Delta(\omega))$ are orthonormal polynomials of degree $i$. The polynomials are chosen such that $X(\omega)$ converges in the $L_2$ sense for any arbitrary stochastic process with finite second moment. This applies to most physical systems.
where $\omega$ is the random event and $\phi_i(\Delta(\omega))$ denotes the gPC basis function of degree $i$, in terms of the random variables $\Delta(\omega)$. For random variables $\Delta$ with certain distributions, the family of orthogonal basis functions $\{\phi_i\}$ can be chosen in such a way such that its weight function $p(\Delta)$ has the same form as the probability density function. With this interpretation,

$$
\int_{\Delta} \phi_i \phi_j p(\Delta) d\Delta = E[\phi_i \phi_j] = E[\phi_i^2] \delta_{ij},
$$

where $E[\cdot]$ denotes the expectation with respect to the probability measure $d\mathcal{M}(\Delta(\omega)) = p(\Delta(\omega)) d\Delta(\omega)$ and probability density function $p(\Delta(\omega))$, and $\Delta$ is the domain of the random variable $\Delta(\omega)$. Henceforth, we will use $\Delta$ to represent $\Delta(\omega)$.

The orthogonal polynomials that are chosen are the members of the Askey-scheme of polynomials [14], which forms a complete basis in the Hilbert space determined by their corresponding support. Table 1 summarizes the correspondence between the choice of polynomials for a given distribution of $\Delta$ [13].

<table>
<thead>
<tr>
<th>Distribution of $p(\Delta)$</th>
<th>$\phi_i(\Delta)$ of the Wiener-Askey Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Hermite</td>
</tr>
<tr>
<td>Uniform</td>
<td>Legendre</td>
</tr>
<tr>
<td>Gamma</td>
<td>Laguerre</td>
</tr>
<tr>
<td>Beta</td>
<td>Jacobi</td>
</tr>
</tbody>
</table>

Table 1 Correspondence between choice of polynomials and given distribution of $\Delta$ [13].

### 2.1 Approximate Solution of Stochastic Differential Equations

A stochastic dynamical system of the form $\dot{x} = f(x, \Delta)$, where $x \in \mathbb{R}^n$ and uncertain parameter $\Delta \in \mathbb{R}^d$, can be solved using the polynomial chaos framework in the following manner. Assume solution of the stochastic differential equation to be $x(t, \Delta)$. For second order processes, the solution for every component of $x \in \mathbb{R}^n$ can be approximated as

$$
\dot{x}_i(t, \Delta) = \sum_{j=0}^{N} x_{ij}(t) \phi_j(\Delta); \ i = 1, \ldots, n.
$$

The series is truncated after $N + 1$ terms, which is determined by the dimension $d$ of $\Delta$ and the order $r$ of the orthogonal polynomials $\{\phi_j\}$, satisfying $N + 1 = \frac{(d+r)!}{d!r!}$. This expression gives the number of terms in a sequence of multi-variate polynomials up to order $r$ with $d$ variables.

Substituting the approximate solution into the dynamical system results in errors

$$
e_i = \dot{x}_i - f_i(\dot{x}, \Delta); \ i = 1, \ldots, n.
$$

The approximation in eqn.(3) is optimal in the $L_2$ sense when the projections of the error on the orthogonal basis functions are zero, i.e.,

$$
\langle e_i(t, \Delta), \phi_j(\Delta) \rangle = \int_{\Delta} e_i(t, \Delta) \phi_j(\Delta) p(\Delta) d\Delta = 0,
$$

for $j = 0, \ldots, N$; $i = 1, \ldots, n$. Equation (4) results in the following $n(N + 1)$ deterministic ordinary differential equations

$$
\dot{x}_{ik} = \frac{\int_{\Delta} f \left( \sum_{j=0}^{N} x_{ij}(t) \phi_j(\Delta) \right) \phi_k(\Delta) p(\Delta) d\Delta}{\int_{\Delta} \phi_k^2(\Delta) p(\Delta) d\Delta},
$$

for $i = 1, \ldots, n$ and $k = 0, \ldots, N$. Therefore, the stochastic dynamics in $\mathbb{R}^n$ has been transformed into deterministic dynamics in $\mathbb{R}^{n(N+1)}$. Let us represent $X_{pc} = [x_{10} \cdots x_{1N} x_{20} \cdots x_{2N} \cdots x_{n0} \cdots x_{nN}]^T$. Then eqn.(5) can be written in a compact form as

$$
\dot{X}_{pc} = F_{pc}(X_{pc}),
$$

where $F_{pc}(X_{pc})$ represents the right hand side of eqn.(5). Equation (6) can be solved using algorithms for ordinary differential equation, to obtain the approximate stochastic response of the system under consideration.
3 Nonlinear Estimation Using Higher Order Moment Updates

Let us consider a nonlinear dynamical system with states $x \in \mathbb{R}^n$ and measured outputs $\tilde{y} \in \mathbb{R}^m$, whose dynamics is governed by

$$\dot{x} = f(x, \Delta),$$

$$\tilde{y} = h(x) + \nu,$$  \hspace{1cm} (7a)

$$\tilde{y} = h(x) + \nu,$$  \hspace{1cm} (7b)

where $\nu$ is the measurement noise with $E[\nu] = 0$ and $E[\nu \nu^T] = R$. The random parameters can be written as $\Delta = [\Delta_{x_0} \Delta_p]_r^T$, where $\Delta_{x_0}$ represents initial condition uncertainty and $\Delta_p$ represents uncertainty in system parameters. Let $p(\Delta)$ be the distribution of $\Delta$.

Estimation algorithms have essentially two steps, the propagation phase and the update phase. It is assumed that $p(\Delta)$ is stationary during the propagation phase. However, the distribution of $x(t, \Delta)$ will not be stationary due to the dynamics. Therefore, the distribution of $\Delta_x$ will change after every update phase. The distribution of $\Delta_p$ will typically not change at all, unless it is updated externally.

Let us also assume that the measurement updates are available at discrete time $t_k, t_{k+1}, \cdots$.

3.1 Step 1: Initialization of State

Given the probability density function of the parameters $p^k(\Delta)$ at time $t_k$, the initial condition for $X_{pc}(t_k)$ at $t_k$ can be obtained using the following equation

$$x_{ij}(t_k) = \int_{D_\Delta} \Delta_{x_{ij}} \phi_i(\Delta)p^k(\Delta)d\Delta,$$  \hspace{1cm} (8)

for $i = 1, \cdots, n$; $j = 0, \cdots, N$. The symbol $\Delta_{x_{ij}}$ represents the $i^{th}$ component of $\Delta_{x_0}$, which is the random variable associated with initial condition uncertainty.

3.2 Step 2: Propagation of Uncertainty & Computation of Prior

With initial condition defined by eqn.(8), the system in eqn.(6) is integrated over the interval $[t_k, t_{k+1}]$ to obtain $X_{pc}(t_{k+1})$, i.e.

$$X_{pc}(t_{k+1}) = X_{pc}(t_k) + \int_{t_k}^{t_{k+1}} F_{pc}(X_{pc}(\tau))d\tau.$$  

The moments of the random process $x(t, \Delta)$ at $t = t_{k+1}$ can be computed from $X_{pc}(t_{k+1})$ as follows (see our previous work [21] for derivations),

$$M_{i}^{1-} = x_{i0},$$  \hspace{1cm} (9)

$$M_{ij}^{2-} = \sum_{p=0}^{N} \sum_{q=0}^{N} x_{ip}x_{jq} \langle \phi_p \phi_q \rangle,$$  \hspace{1cm} (10)

$$M_{ij}^{3-} = \sum_{p=0}^{N} \sum_{q=0}^{N} \sum_{r=0}^{N} x_{ip}x_{jq}x_{kr} \langle \phi_p \phi_q \phi_r \rangle,$$  \hspace{1cm} (11)

$$M_{ijkl}^{4-} = \sum_{p=0}^{N} \sum_{q=0}^{N} \sum_{r=0}^{N} \sum_{s=0}^{N} x_{ip}x_{jq}x_{kr}x_{is} \langle \phi_p \phi_q \phi_r \phi_s \rangle,$$  \hspace{1cm} (12)

and so on; for $i, j, k, l = 1, \cdots, n$. In the above equations $x_{ij} := x_{ij}(t_{k+1})$ and $M_i^{1-}$ represents the prior $i^{th}$ moment at $t_{k+1}$. The inner products of the basis functions are computed with respect to $p^k(\Delta)$, i.e.

$$\langle \phi_p \phi_q \phi_r \phi_s \rangle = \int_{D_\Delta} \phi_p(\Delta)\phi_q(\Delta)\phi_r(\Delta)\phi_s(\Delta)p^k(\Delta)d\Delta.$$
3.3 Step 3: Update Phase

We incorporate the measurements \( \tilde{y} := \tilde{y}(t_k) \) and the prior moments \( M^{i-} \) to get the aposteriori estimates of the moments, \( M^{i+} \). Here we consider the state estimate \( \hat{x}^- \) to be the expected value of \( x(t, \Delta) \) at \( t_{k+1} \), i.e. \( \hat{x}^- = M^{i-} = E[x] \). Also, \( \tilde{y}^- = h(\hat{x}^-) \). Let

\[
\nu = \tilde{y} - h(\hat{x}^-) = h(x) + \nu - h(\hat{x}^-).
\]

Using the approach used by Julier et al.[28,30] and Park et al.[31], we use a linear Kalman gain \( K \) to update the moments. Although updates with nonlinear gains are also possible, but not considered in this paper. Linear update law has also been used by Majji et al.[29] in their design of nonlinear estimators using higher order moment updates. The Kalman gain \( K \) is computed as

\[
K = P^{-1}(x, y)(P^{yv})^{-1}
\]

where \( P^{xv} = E[x, y] \) and \( P^{vuv} = E[v, v] \). This gain is optimal in the minimum variance sense. The update equations for the moments are therefore (see [28,29] for derivations),

\[
\begin{align*}
M^{1+} &= M^{1-} + K
\end{align*}
\]

(14)

\[
\begin{align*}
M^{2+} &= M^{2-} - K P^{xv} K^T
\end{align*}
\]

(15)

\[
\begin{align*}
M^{3+} &= M^{3-} + 3K^2 P^{xuv} - 3K P^{xv} - K^3 P^{vvv}
\end{align*}
\]

(16)

\[
\begin{align*}
M^{4+} &= M^{4-} - 4K P^{xvv} + 6K^2 P^{xuv} - 4K^3 P^{vuv} + K^4 P^{vvv}
\end{align*}
\]

(17)

The fifth and higher order moment update equations can be computed in a similar fashion, using appropriate prior moments and Kalman gain [28,29].

The tensors \( P^{xv} \), \( P^{vuv} \), \( P^{xuv} \), \( P^{xxv} \), \( P^{xvv} \), \( P^{vvv} \), and \( P^{vvv} \) can be computed in terms of the gPC coefficients. Here we only derive the expression for \( P^{xv} \) and \( P^{vuv} \) due to space constraints. Using this derivation process, the higher order tensors can be computed very easily. The expressions for \( P^{xv} \) and \( P^{vuv} \) are

\[
P^{xv} = E[(x - \hat{x})(y - \hat{y})^T] = E[x\hat{y}^T] - \hat{x}E[\hat{y}^T]
\]

\[
= E[x(h(x) + \nu)^T] - \hat{x}E[(h(x) + \nu)^T]
\]

\[
= \int_{\Delta} \left( \sum_i x_i \phi_i(\Delta) \right) h^T \left( \sum_j x_j \phi_j \right) p^k(\Delta) d\Delta - \hat{x} \int_{\Delta} h^T \left( \sum_j x_j \phi_j(\Delta) \right) p^k(\Delta) d\Delta,
\]

where \( x_i = [x_{1i}, x_{2i}, \ldots, x_{ni}]^T \). Similarly,

\[
P^{vuv} = \int_{\Delta} h \left( \sum_j x_j \phi_j(\Delta) \right) h^T \left( \sum_j x_j \phi_j(\Delta) \right) p^k(\Delta) d\Delta + R - \hat{y}\hat{y}^T.
\]

The above expressions for \( P^{xv} \) and \( P^{vuv} \) are dependent on the output function \( h(x) \). If \( h(x) \) is polynomial function, the above expressions become functions of \( M^{i-} \). This is shown in section 5, where we consider the output equation \( h(x) = x^T x \). When \( h(x) \) is a transcendental function, computation of the above integrals cannot be performed directly. One approach would be to expand \( h(x) \) about \( E[x] \) in terms of the perturbations, using Taylor series expansion [32], and obtain a polynomial approximation of \( h(x) \). While Taylor series approximation is straightforward and generally computationally cost effective, it becomes severely inaccurate when higher order gPC expansions are required to represent the physical variability [32]. For example, a 5th order Taylor series approximation using a 3rd order polynomial expansion would require tensor products of six 3rd order basis functions. This will result in 18th order polynomial. This will increase if higher order Taylor series or gPC expansions are used to obtain better approximations. It is well known that computation of higher order polynomials using finite bit representation of real numbers have associated numerical errors. At the same time, for many nonlinear functions this Taylor series approximation is limited by the theoretical range of convergence Taylor series. To tackle the problem of inaccuracies in the evaluation of transcendental functions, using Taylor series expansions, a more robust algorithm is presented by Deusschere et al.[32]. This method is valid for any non polynomial function \( u(\xi) \) for which \( \frac{du}{d\xi} \) can be expressed as a rational function of \( \xi, u(\xi) \). The same issues are encountered with \( f(x, \Delta) \), which defines the dynamics of the system.
3.4 Step 4: Estimation of Posterior Probability Distribution

It is well known that the gPC framework is a method of moments and thus it is difficult to estimate the probability density function of the process, except when Gaussian behaviour is assumed [17]. In this estimation algorithm, the probability density function $p^{k+1}$ at time $t = t_{k+1}$ is determined using maximum entropy estimation theory, subject to constraints defined by the posterior moments $M^{i+}$. This is the solution of the following optimization problem,

$$
\max_{p^{k+1}(\Delta)} \quad -\int_{D_\Delta} p^{k+1}(\Delta) \log(p^{k+1}(\Delta)) d\Delta,
$$

subject to

$$
\int_{D_\Delta} \Delta p^{k+1}(\Delta) d\Delta = M^{1+},
$$

$$
\int_{D_\Delta} Q_2(\Delta)p^{k+1}(\Delta) d\Delta = M^{2+},
$$

$$
\int_{D_\Delta} Q_3(\Delta)p^{k+1}(\Delta) d\Delta = M^{3+},
$$

$$
\int_{D_\Delta} Q_4(\Delta)p^{k+1}(\Delta) d\Delta = M^{4+},
$$

and so on. The symbols $Q_i(\Delta)$ are tensors of polynomials defining the moments corresponding to $M^{i+}$. The functional space of $p^{k+1}(\Delta)$ is approximated using Gaussian mixture models (GMM). In general, any parameterization of $p^{k+1}(\Delta)$ is possible, which will be explored in our future work. Under GMM approximation, $p^{k+1}(\Delta)$ is parameterized as

$$
p^{k+1}(\Delta) = \sum_{i}^M \alpha_i N(\mu_i, \Sigma_i),
$$

with $\alpha_i \in \mathbb{R}$, $\mu_i \in \mathbb{R}^n$, $\Sigma_i = \Sigma_i^T \in \mathbb{R}^{n \times n}$ and $\Sigma_i \geq 0$. For computational simplicity, we assume $\Sigma_i$ to be diagonal, i.e. $\Sigma_i = diag(\sigma_{i1} \cdots \sigma_{in})$. For $p^{k+1}(\Delta)$ to be a probability density function the following constraints need to be satisfied:

$$
\int_{D_\Delta} p^{k+1}(\Delta)d\Delta = 1 \Rightarrow \sum_{i}^M \alpha_i = 1,
$$

$$
p^{k+1}(\Delta) \geq 0 \Rightarrow \alpha_i \geq 0.
$$

Equation (21) and eqn.(22) together implies

$$
0 \leq \alpha_i \leq 1.
$$

With GMM approximation of $p^{k+1}(\Delta)$, the integrals in eqn.(18) and eqn.(19) can be analytically computed and expressed in terms of the unknowns $\alpha_i$, $\mu_i$ and $\Sigma_i$. For a given term in the summation in eqn.(20), the cost in

Preprint submitted to Automatica
Received February 7, 2010 16:31:23 PST
eqn.(18) and constraints in eqn.(19), for $x \in \mathbb{R}^2$, is given by the following expressions.

\[
J_i = \alpha_i (\log(2) - \log(\alpha_i) + \log(\sigma_{11}) + \log(\sigma_{12}) + \log(\pi) + 1)
\]

\[
C_{i,1} = \mu_i \alpha_i,
\]

\[
C_{i,2} = \mu_i \sigma_{11},
\]

\[
C_{i,3} = \mu_i^2 \alpha_i + \sigma_{11}^2,
\]

\[
C_{i,4} = \mu_i \mu_i \alpha_i,
\]

\[
C_{i,5} = \mu_i^2 \alpha_i + \sigma_{12}^2,
\]

\[
C_{i,6} = \mu_i^3 \alpha_i + 3\mu_i \alpha_i \sigma_{11}^2,
\]

\[
C_{i,7} = \mu_i \mu_i \alpha_i + \mu_i \alpha_i \sigma_{11}^2 + \sigma_{12}^2,
\]

\[
C_{i,8} = \mu_i \mu_i^2 \alpha_i + \mu_i \alpha_i \sigma_{12}^2,
\]

\[
C_{i,9} = \mu_i^3 \alpha_i + 3\mu_i \alpha_i \sigma_{12}^2,
\]

\[
C_{i,10} = \mu_i^4 \alpha_i + 6\mu_i^2 \alpha_i \sigma_{11}^2 + 3\alpha_i \sigma_{11}^4,
\]

\[
C_{i,11} = \mu_i^3 \mu_i \alpha_i + 3\mu_i \alpha_i \sigma_{11}^2,
\]

\[
C_{i,12} = \mu_i^2 \mu_i \alpha_i + \mu_i \alpha_i \sigma_{12}^2 + \mu_i^2 \alpha_i \sigma_{11}^2 + \alpha_i \sigma_{11}^2 \sigma_{12}^2,
\]

\[
C_{i,13} = \mu_i \mu_i \alpha_i + \mu_i \alpha_i \sigma_{12}^2 + \sigma_{12}^4,
\]

\[
C_{i,14} = \mu_i \alpha_i + 6\mu_i \alpha_i \sigma_{12}^2 + 3\alpha_i \sigma_{12}^4.
\]

The expressions in the above equations can also be obtained analytically for $x \in \mathbb{R}^n$. The cost function in eqn.(18) and the left hand side of eqn.(19) are obtained by adding the above expressions over $i = 1, \ldots, M$. It can be seen that the cost function is convex in $\alpha_i$ and concave in $\sigma_{ij}$. The constraints are convex in $\alpha_i$ and $\sigma_{ij}$, and not all the constraints are convex in $\mu_{ij}$. The problem can be made convex by restricting $\mu_{ij} \geq 0$, which will require affine transformation of the state variables and rewriting the dynamics and output equation in terms of the new variables. This step however requires the assumption that $p^{k+1}(\Delta)$ is almost zero beyond a finite subdomain of $D_\Delta$, which may be acceptable in some applications. The optimization problem, as presented here, can also be solved as a nonlinear programming problem.

3.5 Step 5: Generation of Basis Functions

Once $p^{k+1}(\Delta)$ has been estimated using GMM approximations, the basis functions $\{\phi_i(\Delta)\}$ need to be generated so that they are orthogonal with respect to the new probability density function. If $\{\phi_i(\Delta)\}$ are orthogonal, the gPC approximations have exponential convergence and thus are optimal. For any other basis functions, the approximation is worse than optimal. However, for some applications use of the same basis functions for all the time steps may provide acceptable results. We use the Gram-Schmidt procedure to generate the set of basis functions $\{\phi_i(\Delta)\}$ that are orthogonal with respect to $p^{k+1}(\Delta)$. Consequently, all the inner products, in the estimation algorithm presented, need to be recomputed at every time step.

The estimation algorithm proceeds by repeating steps one to five as described above.

4 Nonlinear Estimation Using Bayesian Framework

In section we use the Bayesian framework to develop the nonlinear estimation algorithm using polynomial chaos theory. At time $t_{k+1}$, let $x_{k+1}(\Delta)$ be the random variable associated with the state, let $\hat{x}_{k+1}$ be the state estimate and $\tilde{y}_{k+1}$ be the measurement. The objective is to incorporate $\tilde{y}_{k+1}$ to determine $\hat{x}_{k+1}$, using the classical Bayesian approach (pg. 377, in ref.[33]). In this framework we first calculate $p(\hat{y}_{k+1} | x_{k+1}(\Delta) = \hat{x}_{k+1})$, the conditional probability density function of $\hat{y}_{k+1}$, given the current state estimate $\hat{x}_{k+1}$. This function is the likelihood function. Next, we find the conditional probability density function of the state, given $\hat{y}_{k+1}$, i.e. $p(x_{k+1}(\Delta) | \hat{y}_{k+1})$. The state estimate $\hat{x}_{k+1}$ is then determined using the maximum likelihood, minimum-variance or minimum error criterion from $p(x_{k+1}(\Delta) | \hat{y}_{k+1})$. The algorithm in the context of polynomial chaos is described next.
4.1 Step 1: Initialization of State & Propagation of Uncertainty

The first two steps in the Bayesian estimation algorithm using polynomial chaos theory is the same as the previous algorithm. Given the probability density function \( p_k(\Delta) \), the initial conditions for \( X_{pc}(t_k) \) are determined using eqn.(8). The propagation of uncertainty is achieved by integrating the system in eqn.(6) over the interval \([t_k, t_{k+1}]\) to obtain \( X_{pc}(t_{k+1}) \). With \( X_{pc}(t_{k+1}) \), higher order moments are obtained as described in the previous algorithm. With these moments, the prior probability density function \( p_{k+1}^-(\Delta) \) is determined using the Gaussian mixture model, also as described in the previous algorithm. Polynomial chaos is used only to determine \( p_{k+1}^-(\Delta) \). The following steps are standard steps in Bayesian estimation and have been presented here for completeness.

4.2 Step 2: Calculating the Posterior Probability Density Function

First, the likelihood function \( p(\tilde{y}_{k+1}|x_{k+1}(\Delta) = \hat{x}_{k+1}) \) is constructed assuming Gaussian measurement noise and the sensor model as shown in eqn.(7). It is defined as

\[
p(\tilde{y}_{k+1}|x_{k+1}(\Delta) = \hat{x}_{k+1}) = \frac{1}{\sqrt{(2\pi)^m|R|}} e^{-\frac{1}{2}(\tilde{y}_{k+1} - h(\hat{x}_{k+1}))^T R^{-1}(\tilde{y}_{k+1} - h(\hat{x}_{k+1}))},
\]

where \(|R|\) is the determinant of measurement noise covariance matrix.

The posterior probability density function of the states is determined next. It is given by the density function of the states given the current measurement, \( p(x_{k+1}(\Delta)|\tilde{y}_{k+1}) \). Using the classical Bayes rule we can write it as,

\[
p(x_{k+1}(\Delta)|\tilde{y}_{k+1}) = \frac{p(\tilde{y}_{k+1}|x_{k+1}(\Delta) = \hat{x}_{k+1})p_{k+1}^-(\Delta)}{C},
\]

where \( p_{k+1}^-(\Delta) \) is the prior probability density function and

\[
C = \int_{\Delta} p(\tilde{y}_{k+1}|x_{k+1}(\Delta) = \hat{x}_{k+1})p_{k+1}^-(\Delta)d\Delta.
\]

4.3 Step 3: Getting the State Estimate

Depending on the criterion function for state estimation, we can compute the estimate \( \hat{x}_{k+1} \) from \( p(x_{k+1}(\Delta)|\tilde{y}_{k+1}) \). The commonly used criterion are:

1. Maximize the probability that \( \hat{x}_{k+1} = x_{k+1}(\Delta) \). This translates to

\[
\hat{x}_{k+1} = \text{mode of } p(x_{k+1}(\Delta)|\tilde{y}_{k+1}).
\]

This estimate is called most probable estimate or the unconditional maximum likelihood estimate.

2. Minimum variance estimate, which translates to

\[
\min_{x_{k+1}(\Delta)} \int_{\Delta} ||x_{k+1}(\Delta) - \hat{x}_{k+1}||^2 p(x_{k+1}(\Delta)|\tilde{y}_{k+1})d\Delta.
\]

3. Minimize maximum \( |x_{k+1}(\Delta) - \hat{x}_{k+1}| \), which translates to

\[
\hat{x}_{k+1} = \text{median of } p(x_{k+1}(\Delta)|\tilde{y}_{k+1}).
\]

This corresponds to minimum error estimate.
4.4 Step 4: Regeneration of Basis functions

With $p_{k+1}(\Delta) := p(x_{k+1}(\Delta)|\tilde{y}_{k+1})$, the basis functions $\{\phi_i(\Delta)\}$ need to be generated so that they are orthogonal with respect to the new probability density function. This step is identical to Step 5, in the previous algorithm.

The estimation algorithm proceeds by repeating steps one to four as described above.

5 Example

We apply the estimation algorithm presented here to the classical duffing oscillator. It is a two state system, $x = [x_1, x_2]^T$, whose dynamical equations are given by

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = -x_1 - \frac{1}{4}x_2 - x_1^3.$$  \hspace{1cm} (26)

Using the gPC scheme we can write, $x_1(t, \Delta) = \sum_{i=0}^{N} x_{1i}(t) \phi_i(\Delta)$ and $x_2(t, \Delta) = \sum_{i=0}^{N} x_{2i}(t) \phi_i(\Delta)$, where $\phi_i$s are the gPC basis functions and $x_{1i}, x_{2i}$ are the gPC coefficients. The deterministic equations are obtained by substituting the expansions of $x_1(t, \Delta)$ and $x_2(t, \Delta)$ in eqn.(26), and projecting them on the basis functions, which gives us

$$\dot{x}_{1i} = x_{2i}; \hspace{0.5cm} \forall i = 0, 1, \ldots, N; \hspace{1cm} (27a)$$

$$\dot{x}_{2i} = -x_{1i} - \frac{1}{4}x_{2i} - r_i; \hspace{0.5cm} \forall i = 0, 1, \ldots, N; \hspace{1cm} (27b)$$

where,

$$r_i = \frac{1}{\langle \phi_i^2 \rangle} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} z_{1j} z_{1k} z_{1l} \langle \phi_j \phi_k \phi_l \phi_i \rangle.$$

In this example, we consider initial condition uncertainty in $x$ with Gaussian distribution, i.e. $\Delta x_0 \sim N(\mu, \Sigma)$. We use a scalar measurement model for the duffing oscillator which is given by

$$\tilde{y} = x^T x + \nu,$$  \hspace{1cm} (28)

with $\mathbb{E}[\nu] = 0$ and $\mathbb{E}[\nu \nu^T] = 0.006$.

In this example we update up to third order moments. Therefore, we only need to compute the tensors $P^{xv}$, $P^{\nu}$, $P^{xv}$, $P^{xx}$, $P^{vv}$, which for this system are given by
\[ P^{xv} = \left( M^{4}_{1111} + M^{4}_{1122} \right) - \left( M^{3}_{11} M^{3}_{11} + M^{3}_{11} M^{3}_{22} \right), \]

\[ P^{xx} = R + M^{4}_{1111} + M^{4}_{1122} + 2M^{4}_{1122} - 2 \left( (M^{4}_{11})^2 + (M^{4}_{22})^2 \right) \left( M^{3}_{11} + M^{3}_{22} \right) + \left( (M^{4}_{11})^2 + (M^{4}_{22})^2 \right)^2, \]

\[ P^{xv} = R + \left( M^{4}_{1111} + 2M^{4}_{11222} + M^{5}_{112222} \right) - 2 \left( M^{4}_{1111} + M^{4}_{1122} \right) \left( (M^{4}_{11})^2 + (M^{4}_{22})^2 \right) - \left( M^{4}_{11} \right) \left( M^{4}_{1111} + 2M^{4}_{1122} + M^{4}_{11222} \right), \]

\[ P^{xxv} = \left( M^{4}_{1111} + M^{4}_{1122} + M^{4}_{1112} + M^{4}_{1122} \right) - \left( M^{4}_{1111} + M^{4}_{1122} \right) \left( M^{4}_{11} \right) - \left( M^{4}_{11} \right) \left( M^{4}_{1111} + 2M^{4}_{1122} + M^{4}_{11222} \right), \]

\[ P^{v} = M^{6}_{111111} + 3M^{6}_{111112} + 3M^{6}_{112222} + M^{6}_{222222} - 3 \left( (M^{6}_{11})^2 + (M^{6}_{22})^2 \right) \left( M^{5}_{1111} + 2M^{5}_{1122} + M^{5}_{11222} \right) - 3 \left( (M^{6}_{11})^2 + (M^{6}_{22})^2 \right) R - 3 \left( (M^{6}_{11})^2 + (M^{6}_{22})^2 \right)^2 \left( M^{5}_{11} + M^{5}_{22} \right) - \left( (M^{6}_{11})^2 + (M^{6}_{22})^2 \right)^3. \]

The prior moments \( M^{i-} \), for \( i = 1, \ldots, 6 \), are computed from the gPC coefficients \( X_{pc} \) and the probability density function \( p^{(\Delta)} \). Using the tensors \( P^{xv}, P^{xx}, P^{xxv}, P^{v}, P^{vv}, P^{v} \), we obtain the posterior moments \( M^{i+} \), for \( i = 1, 2, 3 \), and subsequently obtain the estimate of the probability density function for the next time step, i.e., for \( p^{(t+1)}(\Delta) \). In this example, we solved the associate optimization problem as a nonlinear programming problem, using SNOPT [34]. In our future work we will solve this problem in the convex optimization framework.

Figure (1) shows the evolution of initial condition uncertainty for the duffing oscillator. We can observe that the probability density function for the state evolve in a non-Gaussian manner. Clearly, linear Gaussian assumptions do not hold for this system.

Figure (2) and fig.(3) illustrate the performance of the EKF based estimator and the proposed gPC based estimator. The initial condition for the system is taken to be \([2 \ 2]^T\) and the uncertainty is assumed to be Gaussian with \( \mu = [1 \ 1]^T \) and \( \Sigma = diag(1,1) \). Therefore, the initial error in the state estimation is \([1 \ 1]^T\). Figure (2(a)) shows the performance of the EKF based estimator with 0.2 seconds measurement update. We observe that the errors in estimates go to zero rapidly. However, when the measurement update interval is increased to 0.3 seconds, the EKF based estimator performs poorly and the errors do not converge to zero. Figure (3(a)) shows the plots of the estimator when gPC theory is used to propagate the uncertainty and only the first two moments are updated, using standard Kalman update law. We observe that this combination achieves better results than the EKF estimator but is inconsistent as the errors escape the ±3\( \sigma \) bounds. Figure (3(b)) shows the performance of the gPC based estimator with third order moment updates, performed every 0.5 seconds. We observe that errors in the estimates converge to zero rapidly and are within the ±3\( \sigma \) bounds. The EKF based estimator for this case diverges. This highlights the importance of using nonlinear dynamics for uncertainty propagation along with higher order moment updates for solving nonlinear estimation problems.

Figure (4) shows the performance of the gPC based Bayesian estimator for the duffing oscillator described above. These simulations were run with larger initial condition uncertainty and slower update rate (1 second). We observe from the plots that the Bayesian estimator is consistent and performs satisfactorily for this system. The EKF based estimator was not stable for the larger initial condition uncertainty and slower update rate.
6 Conclusions

In this paper we present two nonlinear estimation algorithms that combines generalized polynomial chaos theory with higher order moment updates and Bayesian framework. The estimation algorithm is applied to the duffing oscillator model with initial condition uncertainty. The results from these estimators are compared with those obtained from EKF based estimator. We observe that when measurement updates are not frequent, the estimation algorithms presented here outperform the EKF based estimator. This highlights the advantage of nonlinear estimators over EKF based estimator, for nonlinear systems with sparse measurements.

From a computational standpoint, the Bayesian estimator presented here is superior than the one with higher order moment updates. The difference is due to the tensor calculations required for updating higher order moment updates. The computation complexity for propagating uncertainty using polynomial chaos, although more efficient than Monte-Carlo, grows factorially with the dimension of parameters. Thus this approach is best applicable for nonlinear systems with fewer parameters and slow update rate.

A comparison of the proposed methods with particle filters has not been performed here as it is out of the scope of this paper. It is our future goal to collaborate with experts in particle filters and compare the various methods for nonlinear estimation on some benchmark problems.

7 Acknowledgements

This research work was conducted under NASA funding from NRA NNH06ZEA001N-HYP, under cooperative agreement number NNX07AC44A: Fundamental Aeronautics: Hypersonics Project, Topic 5.4 Advanced Control Methods, Subtopic A.5.4.1 Advanced Adaptive Control, with Dan Moerder from NASA Langley as the technical monitor.

References


Fig. 1. Evolution initial condition uncertainty for duffing oscillator. Probability density functions were obtained using high fidelity Monte-Carlo simulations.
Fig. 2. Performance of EKF estimators.

(a) EKF based estimator with 0.2s update.

(b) EKF based estimator with 0.3s update.

Fig. 2. Performance of EKF estimators.
Fig. 3. Performance of gPC estimators.

(a) gPC based estimator with first two moments updated every 0.3s.

(b) gPC based estimator with first three moments updated every 0.5s.

Fig. 3. Performance of gPC estimators.
Fig. 4. Performance of gPC based Bayesian estimators. Update rate is 1 seconds.

(a) gPC based Bayesian estimator - Most probable estimate.
(b) gPC based Bayesian estimator - Minimum error estimate.
(c) gPC based Bayesian estimator - Minimum variance estimate.