Recursive Maximum Likelihood Parameter Estimation for State Space Systems using Polynomial Chaos Theory [⋆]

Benjamin L. Pence [a], Hosam K. Fathy [b], Jeffrey L. Stein [a]

[a] *Mechanical Engineering Department, The University of Michigan, Ann Arbor, MI 48109-2125*

[b] *Mechanical and Nuclear Engineering Department, Pennsylvania State University, University Park, PA 16802*

Abstract

This paper combines polynomial chaos theory with maximum likelihood estimation for a novel approach to recursive parameter estimation in state-space systems. A simulation study compares the proposed approach with the extended Kalman filter to estimate the value of an unknown damping coefficient of a nonlinear Van der Pol oscillator. The results of the simulation study suggest that the proposed polynomial chaos estimator gives comparable results to the filtering method but may be less sensitive to user-defined tuning parameters. Because this recursive estimator is applicable to linear and nonlinear dynamic systems, the authors portend that this novel formulation will be useful for a broad range of estimation problems.

Key words: Estimation Theory, Identification Methods, Linear/Nonlinear Models, Maximum Likelihood, Polynomial Chaos Theory,

1 Introduction

This paper describes a novel method for recursively estimating the unknown static parameters of linear and nonlinear state space systems. This method combines two established theories: generalized polynomial chaos (gPC) theory [21] and maximum likelihood estimation theory (see pages 542-548 of [10]).

Unlike many traditional methods such as recursive least squares (see pages 192-226 of [5]) and total least squares (see pages 381-389 of [10]), the proposed method does not require the underlying model to be formatted into a regressor model form but can be applied directly to state space models. Other state-space estimation methods, such as Kalman filtering approaches (see Chapters 13–15 of [16]) and sequential Monte Carlo (or particle filtering) (pages 35-62 of [12] and [1]) approaches treat unknown parameters as dynamic states and formally include them in the state vector, thus differing from the proposed approach. This paper will use a numerical simulation to study the benefits of the proposed approach compared with the filtering methods.

Many researchers in the estimation community have recognized the benefit of combining polynomial chaos theory with parameter estimation. Blanchard *et al.* combined polynomial chaos theory with the extended Kalman filter for state and parameter estimation [3]. Li and Xiu proposed a gPC ensemble Kalman filter for improved estimation accuracy and computational efficiency [7]. Saad *et al.* proposed a gPC-based ensemble Kalman filter for system identification and monitoring [13], and Smith *et al.* combined gPC with the Luenberger observer for state estimation [17]. Polynomial chaos theory has also been used for parameter estimation - not combined with a state observer. Blanchard *et al.* proposed a Bayesian parameter estimator that selects estimates based on the maximum *a posteriori* estimate [2]. This estimator calculates parameter estimates in a batch manner after all the data has been acquired. Marzouk and Xiu [9] proposed a Bayesian approach to estimate parameters of systems governed by partial differential equations, and provided a valuable study on the convergence of the polynomial chaos based estimators. They used the stochastic collocation approach and extended earlier but similar work done by Marzouk *et al.* [8] which used the Galerkin method. Finally, Southward developed a framework for recursive parameter estimators based on gPC theory [18]. Southward's method used instantaneous gradients of quadratic cost functions to recursively calculate parameter estimates. Shimp [15]

[⋆] Corresponding author B.L. Pence. Tel. (734) 936-3336.

Email addresses: bpence@umich.edu (Benjamin L. Pence), hkf2@psu.edu (Hosam K. Fathy), stein@umich.edu (Jeffrey L. Stein).

and Pence *et al.* [11] applied Southward's method to the problem of real-time vehicle mass estimation.

This paper combines polynomial chaos theory with maximum likelihood estimation to recursively estimate the static unknown parameters of state space systems. Similar to the approaches discussed above, this paper applies polynomial chaos theory to solve the stochastic differential equations that govern the underlying system dynamics. However, unlike any of the methods above, this paper *recursively* calculates the *maximum likelihood values* of the unknown parameters *based on all of the past system observations*. To the best of the authors' knowledge, this is the first paper to address recursive maximum likelihood parameter estimation for state space systems using polynomial chaos theory.

2   Generalized Polynomial Chaos Theory

The generalized polynomial chaos (gPC) framework is essential to the methods of this paper. The gPC framework was developed by Xiu and Karniadakis [21] building off groundbreaking work by Ghanem and Spanos [4] and the conceptualization by Wiener [19].

A set of continuous-time state equations, which are often nonlinear, are used to describe the dynamic behavior of a system.

$$\dot{x} = f(t, x, u; \theta), \qquad 0 \le t \le t_f \qquad (1)$$

$$x(0) = x_0 \qquad (2)$$

The vector $x \in \mathbb{R}^{n_s}$ contains the system states which have known initial conditions $x_0 \in \mathbb{R}^{n_s}$, and the vector $\theta = [\theta_1 \quad \theta_2 \quad \ldots \quad \theta_{n_p}]^T$ contains the unknown parameters. If any of the initial conditions x_0 is unknown, it can be treated as one of the unknown parameters. The input vector $u \in \mathbb{R}^{n_u}$ is known and time-varying. The "dot" notation signifies the derivative with respect to time t.

In general, observations on a system may be governed by a nonlinear, time-varying output model $y = h(t, x, u; \theta)$. However, the scope of this paper is limited to systems having observations described by a linear, time-invariant, discrete-time model:

$$y_k = C(\theta)x(t_k) + v_k. \qquad (3)$$

The output vector $y_k \in \mathbb{R}^{n_y}$ contains the observations on the system at time t_k. The vector $v_k \in \mathbb{R}^{n_y}$ represents an additive Gaussian disturbance with known covariance $R_k \in \mathbb{R}^{n_y \times n_y}$.

The unknown parameters are viewed as being functions of random variables ξ_i, *i.e.*, $\theta_i = \theta_i(\xi_i)$ for $i = 1, \ldots, n_p$.

The random variables are independently identically distributed (IID), and the joint density function is $\overline{\rho}(\xi) = \Pi_{i=1}^{n_p} \rho(\xi_i)$ where $\rho(\xi_i)$ is the distribution of the ith random variable ξ_i, and $\xi = [\xi_1 \quad \xi_2 \quad \cdots \quad \xi_{n_p}]$. Parametric uncertainty leads to uncertainty in the system states. Therefore, $x(t) = x(t, \xi)$ is also a function of the random variables ξ.

Following the gPC method, the unknown parameters $\theta(\xi)$ and system states $x(t, \xi)$ are expanded in terms of orthogonal polynomial basis functions $\Phi_\alpha(\xi)$:

$$\theta(\xi) \approx \sum_{|\alpha|=0}^{S} \theta_\alpha \Phi_\alpha(\xi), \qquad (4)$$

$$\hat{x}(t, \xi) \approx \sum_{|\alpha|=0}^{S} x_\alpha(t) \Phi_\alpha(\xi). \qquad (5)$$

Here, the vector $\alpha := [\alpha_1, \ldots, \alpha_{n_p}]$ is an n_p-dimensional multi-index, and $|\alpha|$ is the sum of the vector elements, *i.e.* $|\alpha| := \alpha_1 + \cdots + \alpha_{n_p}$. Each element α_i of α can take on a non-negative integer value between 0 and S. Under certain assumptions (see [21]), Equations (4) and (5) become exact in the L^2 sense as $S \to \infty$. An infinite expansion is not computationally attainable, so truncation is neccesary, and (4) and (5) are only approximations.

The expansion coefficients $\theta_\alpha, |\alpha| \le S$ are chosen such that (4) is distributed according to the known parameter prior distribution $\overline{\rho}(\theta)$, and hence θ_α is known for all α. Polynomial chaos theory then solves for the coefficients $x_\alpha(t)$ of the polynomial chaos state expansion (5) using either the Galerkin [4] or collocation approach [20]. Some helpful examples that use the Galerkin and collocation approaches can be found in [4], [14], [7], and [11].

2.1   Galerkin Approach

The Galerkin approach solves for the expansion coefficients $x_\alpha(t)$ by projecting the state equations (1) and (2) onto the polynomial chaos basis functions $\Phi_\alpha(\xi)$, *i.e.*,

$$\langle \dot{\hat{x}}(t, \xi), \Phi_\alpha(\xi) \rangle = \langle f(t, \hat{x}(t, \xi), u(t); \theta(\xi)), \Phi_\alpha(\xi) \rangle,$$
$$0 \le t \le t_f;$$
$$\langle \hat{x}(0), \Phi_\alpha(\xi) \rangle = \langle x_0, \Phi_\alpha(\xi) \rangle, \qquad |\alpha| \le S. \quad (6)$$

This results in a set of deterministic state equations having the state-expansion coefficients $x_\alpha(t)$ as the new state variables. These new deterministic state equations can be solved using numerical integration. The number of states in the new set of deterministic state equations is equal to the total number of state-expansion coefficients multiplied by the number n_s of original states. The total

number of state-expansion coefficients $x_\alpha(t)$ (and polynomial chaos basis functions $\Phi_\alpha(\xi)$) is [21]

$$r := \frac{(S + n_p)!}{S! n_p!}. \qquad (7)$$

This number grows rapidly as the polynomial order S and/or the number of unknown parameters n_p increases. The inner product $\langle F(\xi), G(\xi) \rangle$ is an integral of the product of $F(\xi)$ and $G(\xi)$, integrated over the event space of the random variables ξ:

$$\langle F(\xi), G(\xi) \rangle := \int G(\xi) F(\xi) W(\xi) d\xi. \qquad (8)$$

The weighting function $W(\xi)$ depends on the choice of polynomial basis functions, and is generally equal to the prior distribution $\overline{\rho}(\xi)$ of the random variables ξ [21].

2.2 Collocation Approach

The collocation approach [20] can be more straightforward to implement than the Galerkin method, especially for nonlinear systems [14]. However, it is generally less accurate than the Galerkin method [20]. A set of collocation points (or nodes) $\mu^{(1)}, \ldots, \mu^{(Q)}$ ($Q \geq r$ and $\mu^{(i)} \in \mathbb{R}^{n_p}$) are drawn from the parameter prior distribution $\overline{\rho}(\xi)$. These collocation points are substituted for the random variables ξ in (1) and (2), i.e.,

$$\dot{z}^{(i)} = f(t, z^{(i)}(t), u(t); \theta(\mu^{(i)})), \qquad i = 1, \ldots, Q, \quad (9)$$

$$z^{(i)}(0) = x_0. \qquad (10)$$

Here, $z^{(i)} \approx \Sigma_{|\alpha|=0}^{S} x_\alpha(t) \Phi_\alpha(\mu^{(i)})$ is the ith deterministic state vector. The resulting Q uncoupled sets of state equations (each set having n_s states) can be solved using numerical integration. Stacking the new sets of states $z^{(i)}$ into a matrix $Z \in \mathbb{R}^{Q \times n_s}$ yields:

$$Z := \begin{bmatrix} (z^{(1)})^T \\ \vdots \\ (z^{(Q)})^T \end{bmatrix}. \qquad (11)$$

The transpose of the state-expansion of (5) can be written as the vector-matrix product:

$$\hat{x}(t, \xi)^T = (P(\xi))^T \chi(t),$$

$$\chi(t) = \begin{bmatrix} (x_{|\alpha|=0}(t))^T \\ \vdots \\ (x_{|\alpha|=S}(t))^T \end{bmatrix}. \qquad (12)$$

In this equation, if $\Phi_\alpha(\xi)$ is the kth element of the column vector $P(\xi) \in \mathbb{R}^r$, then $x_\alpha^T(t)$ is the kth row of $\chi(t) \in \mathbb{R}^{r \times n_s}$. The matrix Z from (11) can be written in terms of P and χ as follows:

$$Z = \begin{bmatrix} (P(\mu_1))^T \\ \vdots \\ (P(\mu_Q))^T \end{bmatrix} \chi(t). \qquad (13)$$

The estimates of $\chi(t)$, i.e. the estimates of the state expansion coefficients, are obtained by left-multiplying both sides of (13) by the pseudo-inverse of $[P(\mu_1) | \cdots | P(\mu_Q)]^T$.

Another way to write the state expansion (5) that will be useful for concise notation in the following sections is to stack the columns of χ, i.e. $\chi^{(1)}, \ldots, \chi^{(n_s)}$, into a single column vector. Then Equation (5) can be written as follows:

$$\hat{x}(t, \xi) = \mathbb{P}(\xi) X(t). \qquad (14)$$

Here, $\mathbb{P}(\xi) \in \mathbb{R}^{n_s \times r \cdot n_s}$ and $X(t) \in \mathbb{R}^{r \cdot n_s}$, are defined respectively as

$$\mathbb{P}(\xi) := \begin{bmatrix} P(\xi)^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P(\xi)^T \end{bmatrix}, \qquad (15)$$

$$X(t) := \begin{bmatrix} \chi^{(1)}(t) \\ \vdots \\ \chi^{(n_s)}(t) \end{bmatrix}. \qquad (16)$$

Sandu et al. [14] provide a note on the relationship between the stochastic collocation and stochastic Galerkin methods, and they also suggest methods for implementing the Galerkin method on nonlinear systems.

In summary, the deterministic part of the system is calculated using either the Galerkin or collocation method. Solving the deterministic dynamic equations results in known trajectories of the time dependent part $x_\alpha(t)$ which is then recombined using (5) with the random variable dependent part $\Phi_\alpha(\xi)$ to obtain the complete stochastic solution $\hat{x}(t, \xi)$.

3 Recursive Parameter Estimation

This section derives the recursive parameter update law for estimating the maximum likelihood values of the random variables ξ given the system output observations. The estimates of the unknown parameters $\theta(\xi)$ are then calculated using (4). The derivations and resulting parameter estimators of this section constitute the main contributions of this paper.

This development assumes that the noise in the system output observations is zero mean and Gaussian with known covariance matrix $R_k \in \mathbb{R}^{n_y \times n_y}$. It also assumes that the system observations y_k are mutually independent for all k. It assumes the uncertainty in (1) and (2) is entirely due to the unknown parameters. Finally it assumes that the polynomial chaos approximations in (4) and (5) are exact. (As mentioned above, this last assumption is satisfied as the number of expansion terms goes to infinity [21]. In practice, the expansion must be truncated after a finite number of terms, and thus the parameter estimates via this method will only approximately satisfy the maximum likelihood criterion.) Under these assumptions, the likelihood function becomes (see pages 542-548 of [10]):

$$\mathcal{L}_k(\xi|y_{0:k}) = \prod_{\tau=0}^{k} \rho(y_\tau|\xi)$$

$$\propto exp\{-\frac{1}{2}\sum_{\tau=0}^{k}(y_\tau - \hat{y}_\tau(\xi))^T R_\tau^{-1}(y_\tau - \hat{y}_\tau(\xi))\}. \quad (17)$$

Here, $\mathcal{L}_k(\xi|y_{0:k})$ is the likelihood at time t_k of the unknown parameters ξ conditioned on a matrix $y_{0:k}$ which contains all of the output observations up to the current time t_k. The function $\rho(y_k|\xi)$ is the conditional probability of the observation y_k at time t_k given ξ; and $\hat{y}_k(\xi)$ is the output of the stochastic model, which, using (14) can be written as

$$\hat{y}_k := C(\theta(\xi))\mathbb{P}(\xi)X(t_k). \quad (18)$$

The maximum likelihood estimate $\hat{\xi}$ is the value of ξ that maximizes the likelihood function (17). Equation (17) is maximized when the magnitude of the negative term in the exponent, *i.e.*, the negative log-likelihood, is minimized:

$$J_k(\xi) := \frac{1}{2}\sum_{\tau=0}^{k}(y_\tau - \hat{y}_\tau(\xi))^T R_\tau^{-1}(y_\tau - \hat{y}_\tau(\xi)). \quad (19)$$

Thus the most likely value of ξ at time t_k is $\hat{\xi}_k = argminJ_k(\xi)$. The ability to update $J_k(\xi)$ iteratively is critical to making the approach of this paper recursive. This paper leverages the linearity of the output model (18) and the benefits of polynomial chaos to separate the time and unknown parameter parts of the equation to make this recursion possible. By substituting (13) and (18) into (19) and performing a few algebraic manipulations, Equation (19) can be written as:

$$J_k(\xi) = \frac{1}{2}\sum_{i=1}^{n_y}\sum_{j=1}^{n_y}(D_k^{y^{(i)}y^{(j)}} - 2C^{(i)}\mathbb{P}D_k^{Xy^{(j)}}$$

$$+ C^{(i)}\mathbb{P}D_k^{XX^T}(C^{(j)}\mathbb{P})^T). \quad (20)$$

In (20), D_k^G is defined as $D_k^G := \Sigma_{\tau=0}^{k}[R_\tau^{-1}]^{(i,j)}G_\tau$ where $G_k \in \{y_k^{(i)}y_k^{(j)}, X(t_k)y_k^{(j)}, X(t_k)(X(t_k))^T\}$. The scalar term $[R_k^{-1}]^{(i,j)}$ is the jth element in the ith row of the inverse covariance matrix R_k^{-1}. Also, the scalar $y^{(l)}$ is the lth element of the observation vector y_k and $C^{(l)}$ is the lth row of the output matrix C. Equation (20) can be updated recursively from time t_k to t_{k+1} since $D_{k+1}^G = D_k^G + G_{k+1}$, and G_k is not a function of ξ.

The forgoing discussion has outlined a procedure for determining recursively $J_k(\xi)$, *i.e.* the term in the exponent of the likelihood function (17). The remaining challenge is to determine the value of ξ that minimizes $J_k(\xi)$, thus maximizing the likelihood at each time step. This can be viewed as an optimization problem in which the objective function $J_k(\xi)$ is time-varying. The following sections offer potential solution approaches.

3.1 Solution via Gradient Descent

Proceeding in a manner similar to and indeed inspired by Southward's [18], this paper proposes a gradient based parameter update law.

$$\hat{\xi}_{k+1} = \hat{\xi}_k - \Gamma_k \frac{\partial J_k(\xi)}{\partial \xi}|_{\xi=\hat{\xi}_k} \quad (21)$$

Here Γ_k is a user-specified gain matrix that may vary in time. In static optimization, if Γ_k is the identity matrix, (21) is a steepest descent method. If Γ_k is the inverse Hessian matrix (matrix of second derivatives), then (21) is Newton's method, and if Γ_k is proportional to the Hessian matrix, then (21) is a modified Newton's method. These static optimization concepts may be helpful for selecting Γ_k.

As stated above, this gradient based approach is similar to the approach taken by Southward [18] (and Shimp [15]). There is, however, an important difference. This paper addresses maximum likelihood parameter estimation, and hence uses gradients of an *integrated cost function*, *i.e.* the cost function (19) is a function of all the data up to time t_k. Southward's method uses gradients of an *instantaneous cost function* and does not propose to maximize a likelihood function.

Substituting (20) into (21) and using the fact that the time-dependent parts $D_k^{y^{(i)}y^{(j)}}$, $D_k^{Xy^{(j)}}$, and $D_k^{XX^T}$ can be moved outside of the partial derivatives since they do not depend on the unknown parameters gives the following update law:

$$\hat{\xi}_{k+1} = \hat{\xi}_k - \Gamma_k \sum_{i=1}^{n_y}\sum_{j=1}^{n_y}[\frac{\partial C^{(i)}\mathbb{P}}{\partial \xi}(D_k^{XX^T}(C^{(j)}\mathbb{P})^T$$

$$- D_k^{Xy^{(j)}})]_{\xi=\hat{\xi}_k}. \quad (22)$$

In practice, it may be helpful to normalize (22) by dividing by $(1 + \|D_k^{Xy^{(j)}}\|_2)$. Then the final update law becomes:

$$\hat{\xi}_{k+1} = \hat{\xi}_k - \Gamma_k \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} \left(\frac{\partial C^{(i)} \mathbb{P}}{\partial \xi} \Big|_{\xi = \hat{\xi}_k} \right.$$
$$\left. \cdot \frac{D_k^{XX^T} (C^{(j)} \mathbb{P})^T |_{\xi = \hat{\xi}_k} - D_k^{Xy^{(j)}}}{(1 + \|D_k^{Xy^{(j)}}\|_2)} \right). \quad (23)$$

The normalization scalar $(1 + \|D_k^{Xy^{(j)}}\|_2)^{-1}$ is not a function of the random variables ξ and therefore (23) still seeks to maximize the likelihood (17).

3.2 Solution via Random Search

The gradient descent solution of Section 3.1 does not guarantee that the estimated parameters will *globally* minimize $J_k(\xi)$, but may potentially select parameters satisfying only *local* minima. This subsection proposes a strategy for enabling the parameter update law to escape local minima in order to satisfy the global maximum likelihood criterion.

The function $J_k(\xi)$ can be evaluated at time t_k for any realization of ξ. This enables the following strategy:

Let $\mathbb{N} := \{1, 2, 3, \ldots\}$ be the set of natural numbers. At time t_{k+1}, select $n_r \in \mathbb{N}$ realizations of ξ randomly, and evaluate the cost of each realization using (20). Then compare these costs with the cost of the current parameter estimate $\hat{\xi}_k$, and set the new estimate $\hat{\xi}_{k+1}$ to be the realization with the lowest cost. This random search strategy allows the algorithm to search any point in the entire parameter space and thus escape local minima.

A guided random search policy combines the random search with the gradient search: The cost of the gradient solution from (23) is compared with the costs of the n_r randomly selected realizations as well as the cost of the current estimate $\hat{\xi}_k$. The new estimate $\hat{\xi}_{k+1}$ is chosen to be the realization with the lowest cost.

4 Example: Nonlinear Oscillator

This section uses a simulation study of a nonlinear Van der Pol oscillator to demonstrate the proposed method and to compare it with a (hybrid) Extended Kalman Filter (EKF) (Chapter 13 of [16]). The equations for the Van der Pol oscillator are given in state-space representation as follows (see pages 54-59 of [6]):

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -x_1 - \varepsilon(x_1^2 - 1)x_2 \quad (24)$$
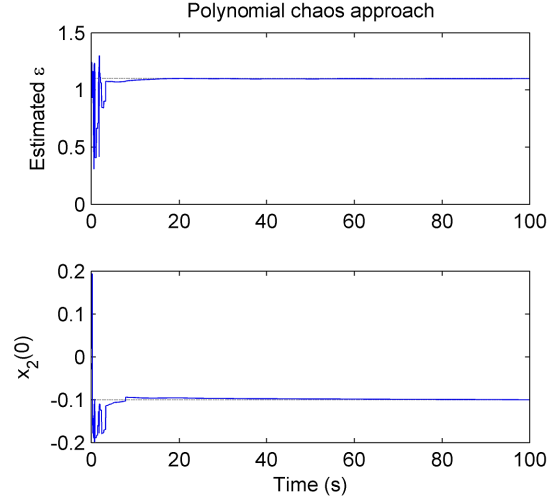$$y_k = x_1(t_k) + v_k.$$

Fig. 1. Convergence of the proposed polynomial chaos estimator. The dotted lines are the true values.

The term $\varepsilon > 0$ represents a nonlinear damping coefficient, and also provides a measure of the nonlinearity of the system. The trajectory of the Van der Pol oscillator tends to a stable limit cycle. The initial conditions were $[x_1(0), x_2(0)] = [0.1, -0.1]$; this study assumes that only $x_1(0)$ is known and that the value of $x_2(0)$ is unknown, but prior information suggests that it could be any value between -0.2 and 0.2 with equal probability. Measurements of the output sequence y_k are sampled at a uniform rate with a step size of 0.01 seconds. The measurement noise v_k is an unknown normally distributed random sequence with constant variance R. Without the added noise, i.e. $v_k = 0$ for all k, the output signal had a mean-squared value of 1.9, and the (assumed unknown) noise variance R was set to be 8 times smaller, i.e., $R = 0.24$. This simulation study assumes that $\varepsilon \sim U[0.3, 1.3]$ is uniformly distributed, and the true (but unknown) value is $\varepsilon = 1.1$.

Because the noise variance R is assumed to be unknown, it becomes a user-specified tuning parameter in both the proposed polynomial chaos algorithm and the EKF filtering algorithm. Because the noise variance is a constant with respect to time and with respect to the unknown parameters, its assumed (positive) value has no effect on the value of ξ that minimizes the cost function shown in (19). It therefore has no effect on the estimate calculated by the proposed polynomial chaos approach. Thus, for all assumed values of R, the convergence of the proposed approach is shown in Figure 1. Figure 2 shows the convergence of the EKF algorithm for different user-assumed values of R. Clearly, the EKF approach is sensitive to the assumed value of R, and hence is more difficult to tune than the polynomial chaos approach for this simulation example.
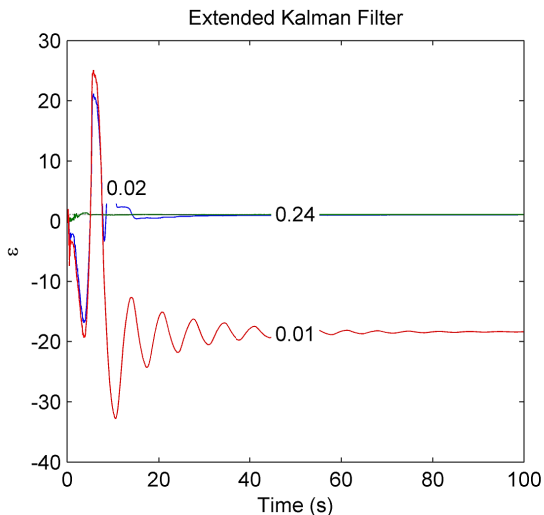
Fig. 2. Convergence of EKF algorithm to various assumed values of $R = 0.01, 0.02, 0.24$. The dotted line is the true value.

5 Conclusions

This paper derived a recursive approach based on polynomial chaos theory for estimating the maximum likelihood values of unknown parameters of state space systems. The derivation assumed that the noise in the system output observations is additive and Gaussian and that the observations are mutually independent. It also assumed the uncertainty in the state equations is due to the unknown parameters. The proposed method was demonstrated using a simulation of a nonlinear Van der Pol oscillator to illustrate the method and to show that it has the potential to be less difficult to tune than the filtering approach. Most importantly, because of its ability to recursively calculate the maximum likelihood values of unknown parameters in both linear and nonlinear systems, the authors believe that the approach will be valuable to a wide range of estimation problems.

Acknowledgements

References

[1] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. Sig. Proc.*, 50:174–188, 2002.

[2] E. Blanchard, A. Sandu, and C. Sandu. A polynomial-chaos-based bayesian approach for estimating uncertain parameters of mechanical systems. *19th Int. Conf. on Des. Theory and Method.; 1st Int. Conf. on Micro- and Nanosys.; and 9th Int. Conf. on Adv. Vehicle Tire Tech., Parts A and B*, 3:1041–1048, 2008.

[3] E. Blanchard, A. Sandu, and C. Sandu. A polynomial chaos-based kalman filter approach for parameter estimation of mechanical systems. *Paper no. 061404, ASME J. of Dynamic Systems Measurement and Control, Special Issue on Physical System Modeling*, 132, 2010.

[4] R. Ghanem and P. Spanos. *Stochastic finite elements: A spectral approach.* Springer-Verlag, New York, 1991.

[5] P.A. Ioannou and J. Sun. *Robust Adaptive Control.* Prentice-Hall, Inc, 1996.

[6] H.K. Khalil. *Nonlinear Systems: Third Ed.* Prentice Hall, Inc., New Jersey, 2002.

[7] J. Li and D. Xiu. A generalized polynomial chaos based ensemble kalman filter with high accuracy. *J. of Comp. Physics*, 228:5454–5694, 2009.

[8] Y.M. Marzouk, N.N. Najm, and L.A. Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *J. Comp. Physics*, 224:560–586, 2007.

[9] Y.M. Marzouk and D. Xiu. A stochastic collocation approach to bayesian inference in inverse problems. *Comm. in Comp. Physics*, 6:826–847, 2009.

[10] T.K. Moon and W.C. Stirling. *Mathematical Methods and Algorithms for Signal Processing.* Prentice-Hall, Inc, New Jersey, 2000.

[11] B.L. Pence, H.K. Fathy, and J.L. Stein. A base-excitation approach to polynomial chaos-based estimation of sprung mass for off-road vehicles. *Proc. ASME Dyn. Sys. and Control Conference 2009*, DSCC2009, n PART A:857–864, 2010.

[12] B. Ristic, S. Maskell, and N. Gordon. *Beyond the Kalman filter: particle filters for tracking applications.* Artech House, Boston, 2004.

[13] G. Saad, R. Ghanem, and S. Masri. Robust system identification of strongly non-linear dynamics using a polynomial chaos based sequential data assimilation technique. *Col. of Tech. Papers, 48th AIAA/ASME/ ASCE/AHS/ASC Struc., Struc. Dyn. and Mat. Conf.*, 6:6005–6013, 2007.

[14] A. Sandu, C. Sandu, and M. Ahmadian. Modeling multibody systems with uncertainties. part 1: theoretical and computational aspects. *Multibody Syst. Dyn.*, 15:373–395, 2006.

[15] S.K. Shimp. Vehicle sprung mass identification using an adaptive polynomial-chaos method. *Masters Thesis*, 2008.

[16] D. Simon. *Optimal State Estimation: Kalman, H infinity, and Nonlinear Approaches.* John Wiley and Sons Inc., New Jersey, 2006.

[17] A.H. Smith, A. Monti, and F. Ponci. Indirect measurements via a polynomial chaos observer. *IEEE Trans. on Instr. and Meas.*, 56:743–752, 2007.

[18] S.C. Southward. Real-time parameter id using polynomial chaos expansions. *Proc. of ASME Int. Mech. Eng. Congress and Expo.*, 9:1167–1174, 2008.

[19] N. Wiener. The homogenous chaos. *Amer. J. Math.*, 60:897–936, 1938.

[20] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys.*, 2:293–309, 2007.

[21] D. Xiu and G.E. Karniadakis. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM J. on Scientific Computing*, 24:619–644, 2002.