Predictive Filtering for Nonlinear Systems

John L. Crassidis[1]

F. Landis Markley[2]

Abstract

In this paper, a real-time predictive filter is derived for nonlinear systems. The major advantage of this new filter over conventional filters is that it provides a method of determining optimal state estimates in the presence of significant error in the assumed (nominal) model. The new real-time nonlinear filter determines ("predicts") the optimal model error trajectory so that the measurement-minus-estimate covariance statistically matches the known measurement-minus-truth covariance. The optimal model error is found by using a one-time step ahead control approach. Also, since the continuous model is used to determine state estimates, the filter avoids discrete state jumps. The predictive filter is used to estimate the position and velocity of nonlinear mass-damper-spring system. Results using this new algorithm indicate that the real-time predictive filter provides accurate estimates in the presence of highly nonlinear dynamics and significant errors in the model parameters.

Introduction

Conventional filter methods, such as the Kalman filter,[1] have proven to be extremely useful in a wide range of applications, including: noise reduction of signals, trajectory tracking of moving objects, and in the control of linear or nonlinear systems. The essential feature of the Kalman filter is the utilization of state-space formulations for the system model. Error in the dynamics system can be separated into "process noise" errors or modeling errors. Process noise errors are usually represented by a zero-mean Gaussian error process with known covariance (e.g., a gyro-error model can be represented by a random walk process). Modeling errors are usually not known explicitly, since system models are not usually improved or updated during the estimation process. The theoretical derivation of the expression for the estimate error

[1]Assistant Professor, *Catholic University of America, Dept. of Mech. Eng., Washington, D.C. 20064.* Member AIAA.

[2] Staff Engineer, *Goddard Space Flight Center, Code 712, Greenbelt, MD 20771*. Associate Fellow AIAA.

covariance in the Kalman filter is only available if one makes assumptions about the model error. The most common assumptions about the model error are that it is also a zero-mean Gaussian noise process. Therefore, in the filter-type literature, most often process noise and model error are treated equally.

The Kalman filter satisfies an optimality criterion which minimizes the trace of the covariance of the estimate error between the system model responses and actual measurements. Statistical properties of the process noise and measurement error are used to determine an "optimal" filter design. Therefore, model characteristics are combined with sequential measurements in order to obtain state estimates which are more accurate than both the measurements and model responses.

As stated previously, errors in the system model of the Kalman filter are usually assumed to be represented by a zero-mean Gaussian noise process with known covariance. In actual practice the noise covariance is usually determined by an *ad hoc* and/or heuristic estimation approach which may result in sub-optimal filter designs. Other applications also determine a steady-state gain directly, which may even produce unstable filter designs.[2] Also, in many cases such as nonlinearities in the actual system responses or non-stationary processes, the assumption of a Gaussian model error process can lead to severely degraded state estimates.

In addition to nonlinear model errors, the actual assumed model may be nonlinear (e.g., three-dimensional kinematic and dynamic equations[3]). The filtering problem for nonlinear systems is considerably more difficult and admits a wider variety of solutions than does the linear problem.[4] The extended Kalman filter is a widely used algorithm for nonlinear estimation and filtering.[5] The essential feature of this algorithm is the utilization of a first-order Taylor series expansion of the model and output system equations. The extended Kalman filter retains the linear calculation of the covariance and gain matrices, and it updates the state estimate using a linear function of the measurement residual; however, it uses the original nonlinear equations for state propagation and in the output system equation.[5] But, the model error statistics are still assumed to be represented by a zero-mean Gaussian noise process.

A new approach for performing optimal state estimation in the presence of significant model error has been developed by Mook and Junkins.[6] This algorithm, called the Minimum Model

Error (MME) estimator, unlike most filter and smoother algorithms, does not assume that the model error is represented by a Gaussian process. Instead, the model error is determined during the MME estimation process. The algorithm determines the corrections added to the assumed model such that the model and corrections yield an accurate representation of the system behavior. This is accomplished by solving system optimality conditions and an output error covariance constraint. Therefore, accurate state estimates can be determined without the use of precise system representations in the assumed model. Also, the MME estimator can be applied to systems with nonlinear models. The MME estimates are determined from a solution of a two-point-boundary-value-problem (see Ref. [6-7]). Therefore, the MME estimator is a batch (off-line) estimator which must utilize post-experiment measurements.

The filter algorithm developed in this paper can be implemented in real-time (as can the Kalman filter). However, the algorithm is not limited to Gaussian noise characteristics for the model error. Essentially, this new algorithm combines the good qualities of both the Kalman filter (i.e., a real-time estimator) and the MME estimator (i.e., determines actual model error trajectories). The new algorithm is based on a predictive tracking scheme first introduced by Lu.[8] Although the problem shown in Ref. [8] is solved from a control standpoint, the algorithm developed in this paper is reformulated as a filter and estimator with a stochastic measurement process. Therefore, the new algorithm is known as a predictive filter. The advantages of the new algorithm include: (i) the model error is assumed unknown and is estimated as part of the solution, (ii) the model error may take any form (even nonlinear), and (iii) the algorithm can be implemented on-line to both filter noisy measurements and estimate state trajectories.

The organization of this paper proceeds as follows. First, the basic equations and concepts used for the filter development are reviewed. Then, a predictive filter is derived for nonlinear systems. This approach determines optimal state estimates in real-time by minimizing a quadratic cost function consisting of a measurement residual term and a model error term. Then, the concept of the covariance constraint is introduced for determining the optimal model error weighting matrix. Finally, an example involving the estimation of the position and velocity in a nonlinear mass-damper-spring system is shown.

Nonlinear Predictive Filter

Preliminaries

In this section, the nonlinear predictive filter algorithm is derived. This development is based upon the duality which exists between the predictive controller for nonlinear systems by Lu[8] and a general estimation problem. In the nonlinear predictive filter it is assumed that the state and output estimates are given by a preliminary model and a to-be-determined model error vector, given by

$$\dot{\hat{\underline{x}}}(t) = \underline{f}\left(\hat{\underline{x}}(t)\right) + G\left(\hat{\underline{x}}(t)\right)\underline{d}(t) \tag{1a}$$

$$\hat{\underline{y}}(t) = \underline{c}\left(\hat{\underline{x}}(t)\right) \tag{1b}$$

where $\underline{f} \in \mathrm{R}^n \to \mathrm{R}^n$ is sufficiently differentiable, $\hat{\underline{x}}(t) \in \mathrm{R}^n$ is the state estimate vector, $\underline{d}(t) \in \mathrm{R}^q$ represents the model error vector, $G\left(\hat{\underline{x}}(t)\right) : \mathrm{R}^n \to \mathrm{R}^{n \times q}$ is the model-error distribution matrix, $\underline{c}\left(\hat{\underline{x}}(t)\right) \in \mathrm{R}^n \to \mathrm{R}^m$ is the measurement vector, and $\hat{\underline{y}}(t) \in \mathrm{R}^m$ is the estimated output vector. State-observable discrete measurements are assumed for Equation (1b) in the following form

$$\tilde{\underline{y}}_k = \underline{c}\left(\underline{x}(t_k)\right) + \underline{v}_k \tag{2}$$

where $\tilde{\underline{y}}_k \in \mathrm{R}^m$ is the measurement vector at time t_k, $\underline{x}(t_k)$ is the true state vector, and $\underline{v}_k \in \mathrm{R}^m$ represents the measurement noise vector which is assumed to be a zero-mean, Gaussian white-noise distributed process with

$$E\left\{\underline{v}_k\right\} = \underline{0} \tag{3a}$$

$$E\left\{\underline{v}_k \underline{v}_l^T\right\} = R\,\delta_{kl} \tag{3b}$$

where $R \in \mathrm{R}^{m \times m}$ is a positive-definite measurement covariance matrix.

A Taylor series expansion of the output estimate in Equation (1b) is given by

$$\hat{\underline{y}}(t + \Delta t) \approx \hat{\underline{y}}(t) + \underline{z}\left(\hat{\underline{x}}(t), \Delta t\right) + \Lambda(\Delta t)\, S\left(\hat{\underline{x}}(t)\right)\underline{d}(t) \tag{4}$$

where the i^{th} element of $\underline{z}(\hat{\underline{x}}(t),\Delta t)$ is given by

$$z_i(\hat{\underline{x}}(t),\Delta t)=\sum_{k=1}^{p_i}\frac{\Delta t^k}{k!}L_f^k(c_i) \tag{5}$$

where p_i, $i=1,2,\ldots,m$, is the lowest order of the derivative of $c_i(\hat{\underline{x}}(t))$ in which any component of $\underline{d}(t)$ first appears due to successive differentiation and substitution for $\hat{x}_i(t)$ on the right side. $L_f^k(c_i)$ is a k^{th} order Lie derivative, defined by (see Ref. [9])

$$\begin{aligned} L_f^k(c_i) &= c_i && \text{for } k=0 \\ L_f^k(c_i) &= \frac{\partial L_f^{k-1}(c_i)}{\partial \hat{\underline{x}}}\underline{f} && \text{for } k\ge 1 \end{aligned} \tag{6}$$

$\Lambda(\Delta t)\in \mathrm{R}^{m\times m}$ is a diagonal matrix with elements given by

$$\lambda_{ii}=\frac{\Delta t^{p_i}}{p_i!},\quad i=1,2,\ldots,m \tag{7}$$

$S(\hat{\underline{x}}(t))\in \mathrm{R}^{m\times q}$ is a matrix with each i^{th} row given by

$$s_i=\left\{L_{g_1}\left[L_f^{p_i-1}(c_i)\right],\ldots,L_{g_q}\left[L_f^{p_i-1}(c_i)\right]\right\},\quad i=1,2,\ldots,m \tag{8}$$

where the Lie derivative with respect to L_{g_j} in Equation (8) is defined by

$$L_{g_j}\left[L_f^{p_i-1}(c_i)\right]\equiv\frac{\partial L_f^{p_i-1}(c_i)}{\partial \hat{\underline{x}}}g_j,\quad j=1,2,\ldots,q \tag{9}$$

Equation (8) is in essence a generalized sensitivity matrix for nonlinear systems.

Nonlinear Filtering

A cost functional consisting of the weighted sum square of the measurement-minus-estimate residuals plus the weighted sum square of the model correction term is minimized, given by

$$J(\underline{d}(t))=\frac{1}{2}\left\{\tilde{\underline{y}}(t+\Delta t)-\hat{\underline{y}}(t+\Delta t)\right\}^T R^{-1}\left\{\tilde{\underline{y}}(t+\Delta t)-\hat{\underline{y}}(t+\Delta t)\right\}+\frac{1}{2}\underline{d}^T(t)\,W\,\underline{d}(t) \tag{10}$$

where $W \in \mathrm{R}^{q \times q}$ is positive semidefinite. Also, a constant sampling rate is assumed so that $\underline{\tilde{y}}(t+\Delta t) \equiv \underline{\tilde{y}}_{k+1}$. Substituting Equation (4), and minimizing Equation (10) with respect to $\underline{d}(t)$ leads to the following model error solution

$$\underline{d}(t) = -\left\{\left[\Lambda(\Delta t) S(\underline{\hat{x}})\right]^T R^{-1} \Lambda(\Delta t) S(\underline{\hat{x}}) + W\right\}^{-1} \left[\Lambda(\Delta t) S(\underline{\hat{x}})\right]^T R^{-1} \left[\underline{z}(\underline{\hat{x}}, \Delta t) - \underline{\tilde{y}}(t+\Delta t) + \underline{\hat{y}}(t)\right] \quad (11)$$

By using the matrix inversion lemma,[10] the model error in Equation (11) can be re-written as

$$\underline{d}(t) = -M(t)\left[\underline{z}(\underline{\hat{x}}, \Delta t) - \underline{\tilde{y}}(t+\Delta t) + \underline{\hat{y}}(t)\right] \quad (12)$$

where

$$M(t) = W^{-1}\left(I - \left[\Lambda(\Delta t) S(\underline{\hat{x}})\right]^T \left\{\Lambda(\Delta t) S(\underline{\hat{x}}) W^{-1} \left[\Lambda(\Delta t) S(\underline{\hat{x}})\right]^T + R\right\}^{-1} \left[\Lambda(\Delta t) S(\underline{\hat{x}})\right] \; W^{-1}\right) \times \left[\Lambda(\Delta t) S(\underline{\hat{x}})\right]^T R^{-1} \quad (13)$$

This form will later be used to show the relationship of the predictive filter to a linear estimator for linear systems. Equation (12) is used in Equation (1a) to perform a nonlinear propagation of the state estimates to time t_k, then the measurement is processed at time t_{k+1} to find the new $\underline{d}(t)$ in $[t_k, t_{k+1}]$, and then the state estimates are propagated to time t_{k+1}. The matrix W serves to weight the amount of model error added to correct the assumed model in Equation (1). As W decreases, more model error is added to correct the model, so that the estimates more closely follow the measurements. As W increases, less model error is added, so that the estimates more closely follow the propagated model.

Covariance Constraint

The weighting matrix (W) in Equation (11) can be determined on the basis that the measurement-minus-estimate error covariance matrix must match the measurement-minus-truth error covariance matrix (see Ref. [6]). This condition is referred to as the "covariance constraint," shown as

$$\frac{1}{N} \sum_{k=0}^{N} \left\{\underline{e}_k - \underline{\bar{e}}\right\}\left\{\underline{e}_k - \underline{\bar{e}}\right\}^T \approx R \quad (14)$$

where $\underline{e}_k \equiv \underline{\tilde{y}}_k - \underline{\hat{y}}_k$, $\underline{\bar{e}}$ is the sample mean of $\underline{\tilde{y}} - \underline{\hat{y}}$, and N is a large number. A test for whiteness can be based upon the autocorrelation function matrix of the measurement residual.[5] The maximum likelihood estimate of the $m \times m$ autocorrelation function matrix for N samples is given by

$$C_k = \frac{1}{N} \sum_{i=k}^{N} \underline{e}_i \, \underline{e}_{i-k}^T \tag{15}$$

A 95% confidence interval for whiteness using a finite sample length is given by[5]

$$\left| \rho_{ii\,k} \right| \le 1.96 / N^{1/2} \tag{16}$$

where ρ_{ii} corresponds to the diagonal elements resulting by normalizing the autocorrelation matrix by the zero-lag elements, given by

$$\rho_{ii\,k} = \frac{c_{ii_k}}{c_{ii_0}} \tag{17}$$

If the confidence interval in Equation (16) and the covariance constraint in Equation (14) are met, then the weighting matrix is optimal. Therefore, the proper balance between model error and measurement residual has been achieved. If the measurement residual covariance is higher than the known measurement error covariance (R), then W should be decreased to less penalize the model error. Conversely, if the residual covariance is lower than the known covariance, then W should be increased so that less unmodeled dynamics are added to the assumed system model.

The sample measurement covariance can be determined from a recursive relationship given by (see Ref. [11])

$$\hat{R}_{k+1} = \hat{R}_k + \frac{1}{k+1} \left[\frac{k}{k+1} \left(\underline{e}_{k+1} - \underline{\bar{e}}_k \right) \left(\underline{e}_{k+1} - \underline{\bar{e}}_k \right)^T - \hat{R}_k \right] \tag{18a}$$

$$\underline{\bar{e}}_{k+1} = \underline{\bar{e}}_k + \frac{1}{k+1} \left(\underline{e}_{k+1} - \underline{\bar{e}}_k \right) \tag{18b}$$

The covariance constraint is met when $\hat{R}_k \to R$, after the filter has converged (i.e., the estimate reaches a stochastic steady-state so that the effects of transients become negligible).

Even though the model error is determined by Equation (11) or (12), it still involves stochastic processes. Therefore, a covariance of the model error can be derived. First, the covariance constraint is re-written as

$$E\left\{\left(\underline{\tilde{y}}_k-\underline{\hat{y}}_k\right)\left(\underline{\tilde{y}}_k-\underline{\hat{y}}_k\right)^T\right\}=R \tag{19}$$

Substituting Equation (2) into Equation (19), and using

$$E\left\{\underline{y}_k\,\underline{v}_k^T\right\}=E\left\{\underline{v}_k\,\underline{y}_k^T\right\}=E\left\{\underline{\hat{y}}_k\,\underline{v}_k^T\right\}=E\left\{\underline{v}_k\,\underline{\hat{y}}_k^T\right\}=0 \tag{20}$$

leads to

$$E\left\{\underline{\tilde{y}}_k\,\underline{\tilde{y}}_k^T\right\}-\underline{\hat{y}}_k\,\underline{\hat{y}}_k^T=R \tag{21}$$

If Equation (14) is satisfied at steady-state, then the following equation is also true

$$E\left\{\underline{\tilde{y}}_{k+1}\underline{\tilde{y}}_{k+1}^T\right\}-\underline{\hat{y}}_{k+1}\underline{\hat{y}}_{k+1}^T=R \tag{22}$$

For a constant sampling interval, Equation (22) is equivalent to

$$E\left\{\underline{\tilde{y}}(t+\Delta t)\underline{\tilde{y}}^T(t+\Delta t)\right\}=\underline{\hat{y}}(t+\Delta t)\underline{\hat{y}}^T(t+\Delta t)+R \tag{23}$$

As long as the process remains stationary, Equation (23) is valid even if the covariance constraint is not satisfied. Also, since the optimal model error solution in Equation (11) is a function of the stochastic measurement noise process, a test for the whiteness of the "determined" model error can be found by using the correlation function in Equations (15)-(17), replacing $\underline{e}$ with $\underline{d}$. If the model error is sufficiently white, then the covariance of the model error can also be determined using a recursive formula shown in Equation (18), again replacing $\underline{e}$ with $\underline{d}$. Another form for the model error covariance can be determined by using Equation (11), and assuming that

$$E\left\{\underline{\hat{y}}(t)\underline{v}^T(t+\Delta t)\right\}=E\left\{\underline{v}(t+\Delta t)\underline{\hat{y}}^T(t)\right\}=0 \tag{24a}$$

$$E\left\{\underline{z}(\underline{\hat{x}},\Delta t)\underline{v}^T(t+\Delta t)\right\}=E\left\{\underline{v}(t+\Delta t)\underline{z}^T(\underline{\hat{x}},\Delta t)\right\}=0 \tag{24b}$$

which leads to

$$E\left\{\underline{d}(t)\,\underline{d}^T(t)\right\} = M(t)\left\{\left\langle \underline{\hat{y}}(t) - \underline{\hat{y}}(t+\Delta t) + \underline{z}(\underline{\hat{x}},\Delta t)\right\rangle + R\right\} M^T(t) \tag{25}$$

where

$$\langle \underline{a} \rangle \equiv \underline{a}\,\underline{a}^T \qquad \text{for any } \underline{a} \tag{26}$$

Therefore, the relative magnitude of the model error can now be determined. In fact, if the determined model error process is truly white, then the inverse of the weighting matrix (W) can be shown to be the maximum likelihood estimate of the model error. This can be used to determine an adaptive scheme for determining W to satisfy the covariance constraint (which will be reported at a later time).

Stability

Filter Stability

The effect of W on filter stability and bandwidth can be determined by applying a discrete error analysis. The filter residual is given by

$$\underline{e}(t+\Delta t) = \underline{\tilde{y}}(t+\Delta t) - \underline{\hat{y}}(t+\Delta t) \tag{27}$$

Substituting Equation (4) into Equation (27) leads to

$$\underline{e}(t+\Delta t) = \left[I - \Lambda(\Delta t)\,S(\underline{\hat{x}})\,M(t)\right]\underline{\hat{e}}(t+\Delta t) \tag{28}$$

where

$$\underline{\hat{e}}(t+\Delta t) \equiv \underline{\tilde{y}}(t+\Delta t) - \underline{\hat{y}}(t) - \underline{z}(\underline{\hat{x}},\Delta t) \tag{29}$$

which is the predicted measurement residual at $t+\Delta t$ assuming $\underline{d} = \underline{0}$.

If S is square and full rank, then $\Lambda\,S\,M$ is also full rank. As $W \to 0$, then $\Lambda\,S\,M \to I$, and $(I - \Lambda\,S\,M) \to 0$. This approaches a deadbeat response for the filter dynamics. As $W \to \infty$, then $M \to 0$, and $(I - \Lambda\,S\,M) \to I$. This yields a filter response with eigenvalues approaching the unit circle. As long as the covariance matrix is positive, the eigenvalues of the filter will lie within the unit circle. Therefore, the filter remains contractive.[12]

Robustness

In the previous section, the filter stability was shown for the linearized system. In this section, filter robustness and stability is shown for the nonlinear system with unmodeled dynamics. This situation may arise when W is not chosen properly. Lu[13] has shown that the dual control problem achieves input/output linearization, and asymptotic tracking of any given trajectory if $p_i \le 4$. An analysis of the robustness properties in the face of unmodeled dynamics for $p_i = 1$, $W = 0$, and square and nonsingular $S(\underline{\hat{x}})$ has also been shown in Ref. [13]. In this paper, the case of $p_i = 1$, $W \ne 0$, and $S(\underline{\hat{x}}) \in \mathrm{R}^{m\times 3}$, where $m \le 3$ is considered. The continuous output estimate for $p_i = 1$ is given by

$$\underline{\hat{\dot{y}}} = \underline{L}_f(\underline{c}) + S(\underline{\hat{x}})\underline{d} \tag{30}$$

where

$$\underline{L}_f(\underline{c}) \equiv \begin{bmatrix} L_f(c_1) \\ \vdots \\ L_f(c_m) \end{bmatrix} \tag{31}$$

Suppose that the unmodeled errors are introduced into the output estimate by

$$\underline{\hat{\dot{y}}} = \underline{L}_f(\underline{c}) + \Delta\underline{L}_f(\underline{c}) + \left[S(\underline{\hat{x}}) + \Delta S(\underline{\hat{x}})\right]\underline{d} \tag{32}$$

and suppose that $\underline{L}_f(\underline{c})$ and $\Delta\underline{L}_f(\underline{c})$ are bounded by

$$\left\|\underline{L}_f(\underline{c})\right\| \le n_1, \quad \left\|\Delta\underline{L}_f(\underline{c})\right\| \le n_2, \quad \text{for all } x \in X \tag{33}$$

Furthermore, assume that $\Delta S(\underline{\hat{x}})$ is represented by

$$\Delta S(\underline{\hat{x}}) = \delta(\underline{\hat{x}})S(\underline{\hat{x}}) \tag{34}$$

where $\delta(\underline{\hat{x}})$ is a scalar, continuous function with bound given by $-1 < \delta(\underline{\hat{x}}) < n_3$. Assuming that the model errors and measurement errors are isotropic leads to $W = w\,I$, and $R = r\,I$. Then, the matrix inverse in Equation (11) can be written as (suppressing arguments)

$$\left\{[\Lambda S]^T R^{-1} \Lambda S + W\right\}^{-1} = \left\{(v-\sigma) I + C\right\}^{-1} \tag{35}$$

where

$$C \equiv \frac{\Delta t^2}{r} S^T S \tag{36a}$$

$$\sigma = \frac{\Delta t^2}{2r} \operatorname{tr}\left(S^T S\right) \tag{36b}$$

$$v = w + \sigma \tag{36c}$$

By the Cayley-Hamilton theorem, any meromorphic function of C can be expressed as a quadratic in C (see Ref. [14]), yielding

$$\left\{(v-\sigma) I + C\right\}^{-1} = \frac{1}{\gamma}\left(\alpha I + \beta C + C^2\right) \tag{37}$$

where

$$\alpha = v^2 - \sigma^2 + k \tag{38a}$$

$$\beta = -(v + \sigma) \tag{38b}$$

$$\gamma = (v - \sigma)\alpha + \Delta = w\alpha + \Delta \tag{38c}$$

$$k = \operatorname{tr}(\operatorname{adj} C) = \frac{\Delta t^4}{r^2} \operatorname{tr}\left[\operatorname{adj}\left(S^T S\right)\right] \tag{38d}$$

$$\Delta = \frac{\Delta t^6}{r^3} \det\left(S^T S\right) \tag{38e}$$

Therefore, the error dynamics become

$$\underline{\dot{e}} = -\frac{\Delta t}{r\gamma}(1+\delta) Q \underline{e} + \underline{\tilde{\dot{y}}} - \underline{L}_f - \Delta \underline{L}_f + \frac{\Delta t^2}{r\gamma}(1+\delta) Q\left[\underline{L}_f - \underline{\tilde{\dot{y}}}\right] \tag{39}$$

where

$$Q \equiv \left[\alpha \left(S\,S^T \right) + \frac{\Delta t^2}{r} \beta \left(S\,S^T \right)^2 + \frac{\Delta t^4}{r^2} \beta \left(S\,S^T \right)^3 \right] \tag{40}$$

Now, define a Lyapunov function $V = \|\underline{e}\|^2 / 2$. Using the norm inequality,[15] and the fact that (see Appendix)

$$\underline{e}^T Q\, \underline{e} \geq \frac{1}{\|Q^{-1}\|} \underline{e}^T \underline{e} \tag{41}$$

leads to

$$\dot{V} \leq -\frac{2\,\Delta t\,(1+\delta)}{r\,\gamma \|Q^{-1}\|} V + \|\underline{e}^T\| \left\| \tilde{\underline{\dot{y}}} - \underline{L}_f - \Delta \underline{L}_f \right\| + \frac{\Delta t^2 (1+\delta)}{r\,\gamma} \|\underline{e}^T\| \, \|Q\| \left\| \underline{L}_f - \tilde{\underline{\dot{y}}} \right\| \tag{42}$$

Next, using the well known inequality $a\,b \leq z\,a^2 + b^2 / (4\,z)$ for any a, b, and $z > 0$, and defining $\xi \equiv \Delta t\,(1+\delta)/(r\,\gamma)$ yields

$$\dot{V} \leq \left(-\frac{2\,\xi}{\|Q^{-1}\|} + 4\,z \right) V + \frac{\left\| \tilde{\underline{\dot{y}}} - \underline{L}_f - \Delta \underline{L}_f \right\|^2 + \Delta t^2 \xi^2 \, \|Q\|^2 \left\| \underline{L}_f - \tilde{\underline{\dot{y}}} \right\|^2}{4\,z} \tag{43}$$

Substituting $4\,z = \xi / \|Q^{-1}\|$ leads to

$$\dot{V} \leq -\frac{\xi}{\|Q^{-1}\|} V + b \tag{44}$$

where

$$b \equiv \|Q^{-1}\| \left[\frac{1}{\xi} \left\| \tilde{\underline{\dot{y}}} - \underline{L}_f - \Delta \underline{L}_f \right\|^2 + \Delta t^2 \xi \|Q\|^2 \left\| \underline{L}_f - \tilde{\underline{\dot{y}}} \right\|^2 \right] \tag{45}$$

Therefore, Equation (44) can be solved to yield

$$V \leq \left(V_0 - \frac{b \|Q^{-1}\|}{\xi} \right) e^{-\xi\,t / \|Q^{-1}\|} + \frac{b \|Q^{-1}\|}{\xi} \tag{46}$$

where $V_0 \geq b \left\| Q^{-1} \right\| / \xi$ is required to maintain the inequality. Defining the bounds

$$\theta = \inf(1+\delta) > 0 \tag{47a}$$

$$\left\| \tilde{\underline{y}} \right\|_\infty = \max_{t \in [0, t_f]} \sum_{i=1}^{m} \left| \tilde{\underline{y}}_i \right| \tag{47b}$$

and using the matrix norm inequality again leads to

$$\left\| \underline{e} \right\| \leq \sqrt{2} \frac{r \gamma \left\| Q^{-1} \right\|}{\Delta t\, \theta} \mu_1^2 + \sqrt{2}\, \Delta t \left\| Q^{-1} \right\| \left\| Q \right\| \mu_2^2 \tag{48}$$

where

$$\mu_1 = \left[\left\| \tilde{\underline{y}} \right\|_\infty - n_1 - n_2 \right] \tag{49a}$$

$$\mu_2 = \left[\theta \left(n_1 - \left\| \tilde{\underline{y}} \right\|_\infty \right) \right] \tag{49b}$$

Assuming that $\left\| S\, S^T \right\| \leq s_1$ for all $S\, S^T$ leads to the following bound on $\left\| Q \right\|$

$$\left\| Q \right\| \leq \alpha\, s_1 + \frac{\Delta t^2}{r} \left| \beta \right| s_1^2 + \frac{\Delta t^4}{r^2} s_1^3 \tag{50}$$

Also, $\mathrm{tr}\left(S^T S \right) \leq 3 s_1$ and $\det\left(S^T S \right) \leq s_1^3$, which leads to

$$\sigma \leq \frac{3 \Delta t^2}{2 r} s_1 \tag{51a}$$

$$k \leq \frac{9\, \Delta t^4}{r^2} s_1^2 \tag{51b}$$

$$\Delta \leq \frac{\Delta t^6}{r^3} s_1^3 \tag{51c}$$

Substituting Equation (51) into Equations (38) and (50) leads to the following bounds on $\left\| Q \right\|$ and γ

$$\|Q\| \le w^2 s_1 + \frac{4\,w\,\Delta t^2 s_1^2}{r} + \frac{13\,\Delta t^4 s_1^3}{r^2} \tag{52a}$$

$$\gamma \le w^3 + \frac{3 w^2 \Delta t^2 s_1}{r} + \frac{9\,\Delta t^4 s_1^2}{r^2} + \frac{\Delta t^6 s_1^3}{r^3} \tag{52b}$$

A bound on $\left\|Q^{-1}\right\|$ is found by writing it as

$$\begin{aligned} \left\|Q^{-1}\right\| &= \gamma \left\| \left\{ S \left[\frac{\Delta t^2}{r} S^T S + wI \right]^{-1} S^T \right\}^{-1} \right\| \\ &\le \gamma \left\| \left(S\,S^T\right)^{-1} \right\| \left\| \left[\frac{\Delta t^2}{r} S^T S + wI \right] \right\| \\ &\le \gamma \left\| \left(S\,S^T\right)^{-1} \right\| \left\{ \frac{\Delta t^2}{r} \left\|S^T S\right\| + w \right\} \end{aligned} \tag{53}$$

Assuming that $\left\|\left(S\,S^T\right)^{-1}\right\| \le s_2$ for all $S\,S^T$ leads to

$$\left\|Q^{-1}\right\| \le \gamma\, s_2 \left[\frac{\Delta t^2}{r} s_1 + w \right] \tag{54}$$

Therefore Equations (48), (52), and (54) define the bound for the error dynamics under unmodeled uncertainty. Similar results can be obtained for $1 < p_i \le 4$. The case where $q > 3$ can also be determined using a Cayley-Hamilton expansion, but becomes increasingly more complicated.

Numerical Stability

If the system is unobservable then $S^T S$ is not full rank. However, the filter can compensate for this by adding more model correction. It can be shown that the filter remains for bounded model uncertainties as long as

$$(v - \sigma)\alpha + \Delta > 0 \tag{55}$$

If $S^T S$ is not full rank, then $\Delta = 0$, which leads to the following condition

$$v > \frac{\Delta t^2}{2r}\,\mathrm{tr}\left(S^T S\right) \tag{56}$$

Therefore, the filter remains contractive as long as $w > 0$. This condition is always met, but Equation (56) can be used to help determine any numerical difficulties (i.e., large values of σ may produce numerical difficulties). One possible solution is to make r as large as possible. However, then w will be adjusted to meet the covariance constraint, so that the numerical difficulties remain. Another solution to this problem is to use smaller sampling interval, but this may not be possible. A more practical solution is to utilize a "$U-D$" factorization of Equation (13) (see Ref. [16]).

Cases

Case 1. Let $p_i = 1$ for both the state and output systems. Equations (5), (7) and (8) reduce to

$$\underline{z} = \Delta t\, H(\hat{\underline{x}})\, \underline{f}(\hat{\underline{x}}) \tag{57a}$$

$$H(\hat{\underline{x}}) \equiv \frac{\partial \hat{\underline{y}}}{\partial \hat{\underline{x}}} \tag{57b}$$

$$\Lambda = \Delta t\, I \tag{57c}$$

$$S = H(\hat{\underline{x}}) G(\hat{\underline{x}}) \tag{57d}$$

Therefore, the model error trajectory in Equation (12) is given by

$$\begin{aligned}\underline{d} = -\Delta t &\left\{ I - W^{-1} G^T H^T \left[H G W^{-1} G^T H^T + \Delta t^{-2} R \right]^{-1} H G \right\} \\ &\times W^{-1} G^T H^T R^{-1} \left\{ \hat{\underline{y}}(t) - \tilde{\underline{y}}(t+\Delta t) + \Delta t\, H\, \underline{f} \right\}\end{aligned} \tag{58}$$

Equations (57-58) can be used to develop a predictive filter for a linear system, given by

$$\dot{\hat{\underline{x}}} = F\, \hat{\underline{x}} + G\, \underline{d} \tag{59a}$$

$$\hat{\underline{y}} = H\, \hat{\underline{x}} \tag{59b}$$

For the linear case Equation (58) is similar to a linear estimator. This can be shown be converting Equation (59) into discrete form

$$\underline{\hat{x}}_{k+1} = \Phi\,\underline{\hat{x}}_k + \Gamma\,\underline{d}_k \tag{60}$$

If the sampling interval in the discrete conversion in Equation (60) is equal to the measurement sampling interval (Δt), and if the first-order approximations of $\Phi \approx [I + \Delta t\, A]$, and $\Gamma \approx \Delta t\, G$ are made, then the following equation for $\underline{d}_k$ is given

$$\underline{d}_k = \left\{ I - W^{-1}\Gamma^T H^T \left[H\,\Gamma W^{-1}\Gamma^T H^T + R \right]^{-1} H\,\Gamma \right\} W^{-1}\Gamma^T H^T R^{-1} \left\{ \underline{\tilde{y}}_{k+1} + H\,\Phi\,\underline{\hat{x}}_k \right\} \tag{61}$$

Case 2. Consider the following system

$$\underline{\dot{\hat{x}}}_1 = \underline{f}_1(\underline{\hat{x}}_1, \underline{\hat{x}}_2) \tag{62a}$$

$$\underline{\dot{\hat{x}}}_2 = \underline{f}_2(\underline{\hat{x}}_2) + G_2(\underline{\hat{x}}_2)\underline{d} \tag{62b}$$

$$\underline{\hat{y}} = \underline{c}(\underline{\hat{x}}_1) \tag{62c}$$

with $p_i = 2$. Equation (62a) usually defines the kinematics, and Equation (62b) usually defines the dynamics of a system. Equations (5), (7) and (8) now become

$$\underline{z} = \Delta t\, \underline{L}_f^1 + \frac{\Delta t^2}{2}\left[\frac{\partial \underline{L}_f^1}{\partial \underline{\hat{x}}_1} \underline{f}_1(\underline{\hat{x}}_1, \underline{\hat{x}}_2) + \frac{\partial \underline{L}_f^1}{\partial \underline{\hat{x}}_2} \underline{f}_2(\underline{\hat{x}}_2) \right] \tag{63a}$$

$$\underline{L}_f^1 \equiv \frac{\partial \underline{c}}{\partial \underline{\hat{x}}_1} \underline{f}_1(\underline{\hat{x}}_1, \underline{\hat{x}}_2) + \frac{\partial \underline{c}}{\partial \underline{\hat{x}}_2} \underline{f}_2(\underline{\hat{x}}_2) \tag{63b}$$

$$\Lambda = \frac{\Delta t^2}{2} I \tag{63c}$$

$$S = \frac{\partial \underline{L}_f^1}{\partial \underline{\hat{x}}_2} G_2(\underline{\hat{x}}_2) \tag{63d}$$

Example

In this section, a simple example which illustrates the application of the predictive filter to a nonlinear mass-damper-spring system is shown. Consider the following system[17]

$$m\ddot{x} + b\,\dot{x}|\dot{x}| + k_1\,x + k_2\,x^3 = 0 \tag{64}$$

where $b\,\dot{x}|\dot{x}|$ represents the nonlinear damping, and $k_1\,x + k_2\,x^3$ represents a linear spring with a nonlinear hardening effect. This system can be shown to be asymptotically stable by choosing the following Lyapunov function[17]

$$V(x) = \frac{1}{2} m\,\dot{x}^2 + \frac{1}{2} k_1\,x^2 + \frac{1}{4} k_2\,x^4 \tag{65}$$

which leads to

$$\dot{V}(x) = -b|\dot{x}|^3 \tag{66}$$

Therefore, the mechanical energy of the system converges to zero for any initial condition. Using the predictive filter approach, the system model is modified by the addition of a to-be-determined unmodeled effect. The state-space representation is given by

$$\begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{x}}_2 \end{bmatrix} = \begin{bmatrix} \hat{x}_2 \\ -c_1\,\hat{x}_2\,|\hat{x}_2| - c_2\,\hat{x}_1 - c_3\,\hat{x}_1^3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} d(t) \tag{67}$$

where $c_1 \equiv b/m$, $c_2 \equiv k_1/m$, $c_3 \equiv k_2/m$, and x_1 and x_2 represent position (x) and velocity ($\dot{x}$), respectively. For this system, the model error is represented as an input to the mass-damper-spring system. The measurements are given by

$$\tilde{y}_k = x_1(t_k) + v_k \tag{68}$$

where the variance of v_k is defined as r. The lowest order time-derivative of Equation (68) in which the model error first appears is two. Therefore, the predictive filter equations are given by Equations (62-63), which is Case 2. For this example, the determined model error is given by

$$d = -\frac{2\,\Delta t^3}{\Delta t^4 + 4\,r\,w}\left[\hat{x}_2 - \frac{\Delta t}{2}\left(c_1\,\hat{x}_2\,|\hat{x}_2| + c_2\,\hat{x}_1 + c_3\,\hat{x}_1^3\right) - \frac{1}{\Delta t}\left(\tilde{y}^{\Delta} - \hat{x}_1\right)\right] \tag{69}$$

where $\tilde{y}^{\Delta} \equiv \tilde{y}_{k+1}$. The case where $w = 0$ corresponds to the feedback linearization case, yielding

$$\begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{x}}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2/\Delta t^2 & -2/\Delta t \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2/\Delta t^2 \end{bmatrix} \tilde{y}^{\Delta} \tag{70}$$

The eigenvalues of the state matrix are given by

$$s_{1,2} = \frac{1}{\Delta t}(-1 \pm j) \tag{71}$$

Therefore, the filter's dynamics are dependent only on the sampling interval. Also, this case represents a linear filter on the measurements only, so that no model error correction is added.

The true state history for this example is given using Equation (67) with $d(t) = 0$ for all t, $c_1 = 10$, $c_2 = 0.1$, and $c_3 = 3$, and initial conditions of $x_1(t_0) = -0.1$, and $x_2(t_0) = 0.1$. A plot of the true states with these parameters is shown in Figure 1. Measurements are obtained by using a sampling interval of 0.1 seconds, and the standard deviation of v_k in Equation (68) is 0.0005. Model error is introduced into the system by perturbing c_2 and c_3, which are chosen to be $c_2 = -100$, and $c_3 = 40$. Also, a weighting factor of $w = 0.09$ was determined by satisfying the covariance constraint once the filter reached steady-state. Even though a significant amount of error is present in the assumed model, the predictive filter is able to accurately estimate for the states, as shown by Figure 2. A plot of the actual and determined model error histories is shown in Figure 3. This example shows that the model error for this example cannot be represented by a zero-mean Gaussian process, as is assumed in the Kalman filter. However, the predictive filter is clearly able to correctly determine the actual model error in the system. Finally, the predictive filter is tested for initial conditions errors. For this test, the assumed initial conditions in the filter are set to zero. Figure 4 depicts the filter convergence for this case. The predictive filter is able to converge very quickly (within 0.06 seconds). This example clearly shows that the predictive filter scheme provides robust performance in a nonlinear system for both significant errors in the assumed model and in the initial conditions.

Conclusions

In this paper, a predictive filter was presented for nonlinear systems. Advantages of the new algorithm over the extended Kalman filter include: (i) the model error is assumed unknown and is estimated as part of the solution, (ii) the model error may take any form (even nonlinear), and (iii) the model error is used to propagate a continuous model which avoids discrete jumps in the state estimate. An example of this algorithm was shown which estimated the position and velocity of nonlinear mass-damper-spring system. Results using this new algorithm indicated that the real-time predictive filter provides accurate estimates in the presence of highly nonlinear dynamics and significant errors in the model parameters.

Appendix

In this section, the inequality given by Equation (41) is proved. The vector $\underline{e}$ is first represented by

$$\underline{e} = \sum_{i=1}^{m} \alpha_i \, \underline{u}_i \tag{A1}$$

where $\underline{u}_i$ are the eigenvectors of Q, and α_i are some scalar coefficients. Therefore, the product $Q\underline{e}$ is given by

$$Q\underline{e} = \sum_{i=1}^{m} \alpha_i \, \lambda_i \, \underline{u}_i \tag{A2}$$

where λ_i are the eigenvalues of Q. Using the fact that

$$\|\underline{e}\|^2 = \sum_{i=1}^{m} \alpha_i^2 \tag{A3}$$

leads to the following inequality

$$\lambda_{\min} \sum_{i=1}^{m} \alpha_i^2 \le \underline{e}^T Q \underline{e} \le \lambda_{\max} \sum_{i=1}^{m} \alpha_i^2 \tag{A4}$$

Therefore, using the following identity

$$\lambda_{\min}(Q) = \frac{1}{\lambda_{\max}\left(Q^{-1}\right)} = \frac{1}{\left\|Q^{-1}\right\|} \tag{A5}$$

then the following inequality must hold true

$$\underline{e}^T Q \underline{e} \geq \frac{1}{\left\| Q^{-1} \right\|} \underline{e}^T \underline{e} \tag{A6}$$

Acknowledgments

The first author's work was supported by a National Research Council Postdoctoral Fellowship tenured at NASA-Goddard Space Flight Center. The author greatly appreciates this support. Also, the author wishes to thank Dr. D. Joseph Mook of the State University of New York at Buffalo, and Dr. Ping Lu of Iowa State University for many interesting and helpful discussions.

References

[1]Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, Vol. 82, March 1960, pp. 34-45.

[2]Mason, P.A.C., and Mook, D.J., "Scalar Gain Interpretation of Large Order Filters," *Proceedings of the Flight Mechanics/Estimation Theory Symposium*, NASA-Goddard Space Flight Center, Greenbelt, MD, 1992, pp. 425-439.

[3]Kane, T.R., Likins, P.W., and Levinson, D.A., *Spacecraft Dynamics*, McGraw-Hill, NY, 1983.

[4]Gelb, A., *Applied Optimal Estimation*, MIT Press, MA, 1974.

[5]Stengel, R.F., *Optimal Control and Estimation*, Dover Publications, NY, 1994.

[6]Mook, D.J., and Junkins, J.L., "Minimum Model Error Estimation for Poorly Modeled Dynamic Systems," *Journal of Guidance, Control and Dynamics*, Vol. 11, No. 3, May-June 1988, pp. 256-261.

[7]Crassidis, J.L., Mason, P.A.C., and Mook, D.J., "Riccati Solution for the Minimum Model Error Algorithm," *Journal of Guidance, Control and Dynamics*, Vol. 16, No. 6, Nov.-Dec. 1993, pp. 1181-1183.

[8]Lu, P., "Nonlinear Predictive Controllers for Continuous Systems," *Journal of Guidance, Control and Dynamics*, Vol. 17, No. 3, May-June 1994, pp. 553-560.

[9]Hunt, L.R., Luksic, M., Su, R., "Exact Linearizations of Input-Output Systems," *International Journal of Control*, Vol. 43, No. 1, 1986, pp. 247-255.

[10]Bierman, G.J., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, FL, 1977.

[11]Lewis, F.L., *Optimal Estimation*, John Wiley & Sons, NY, 1986.

[12]Vidyasagar, M., *Nonlinear Systems Analysis*, Prentice Hall, NJ, 1993.

[13]Lu, P., "Optimal Predictive Control of Continuous Nonlinear Systems," *International Journal of Control*, Vol. 62, No. 3, Sept. 1995, pp. 633-649.

[14]Shuster, M.D., and Oh, S.D., "Attitude Determination from Vector Observations," *Journal of Guidance and Control*, Vol. 4, No. 1, Jan.-Feb. 1981, pp. 70-77.

[15]Horn, R.A., and Johnson, C.R., *Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[16]Thornton, C.L., and Jacobson, R.A., "Linear Stochastic Control Using the UDU^T Matrix Factorization," *Journal of Guidance and Control*, Vol. 1, No. 4, July-Aug. 1978, pp. 232-236.

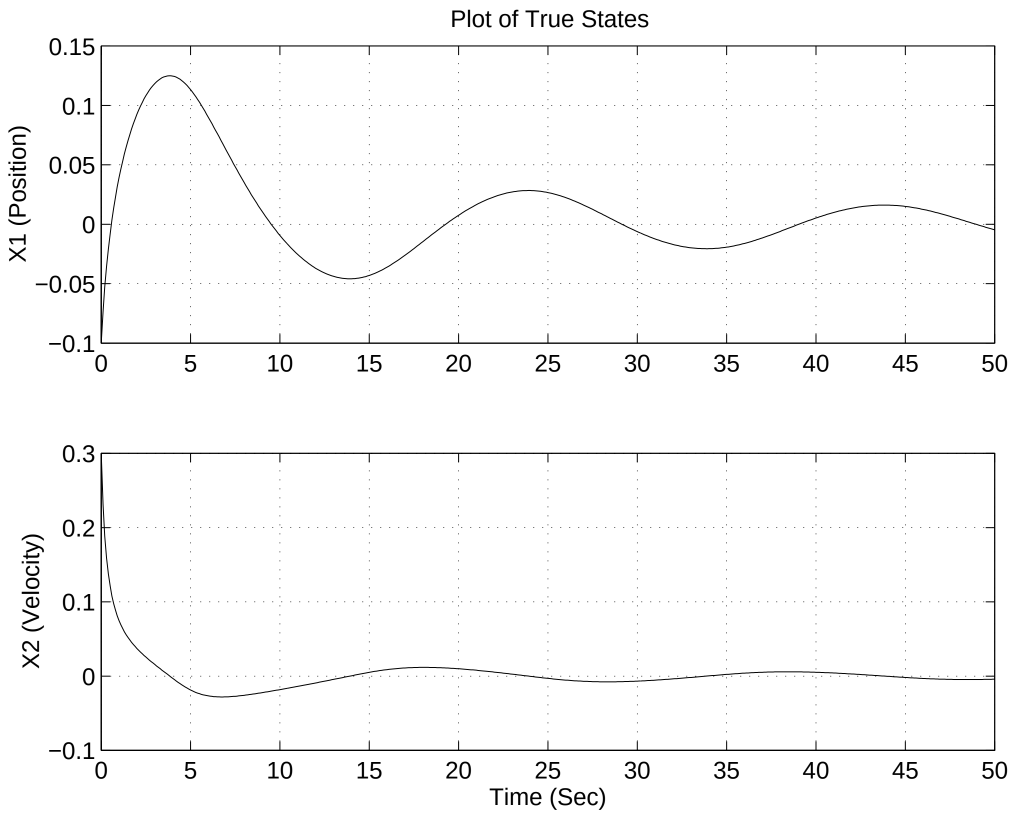
[17]Slotine, J.J.E, and Li, W., *Applied Nonlinear Control*, Prentice Hall, NJ, 1991.

Figure 1 True States

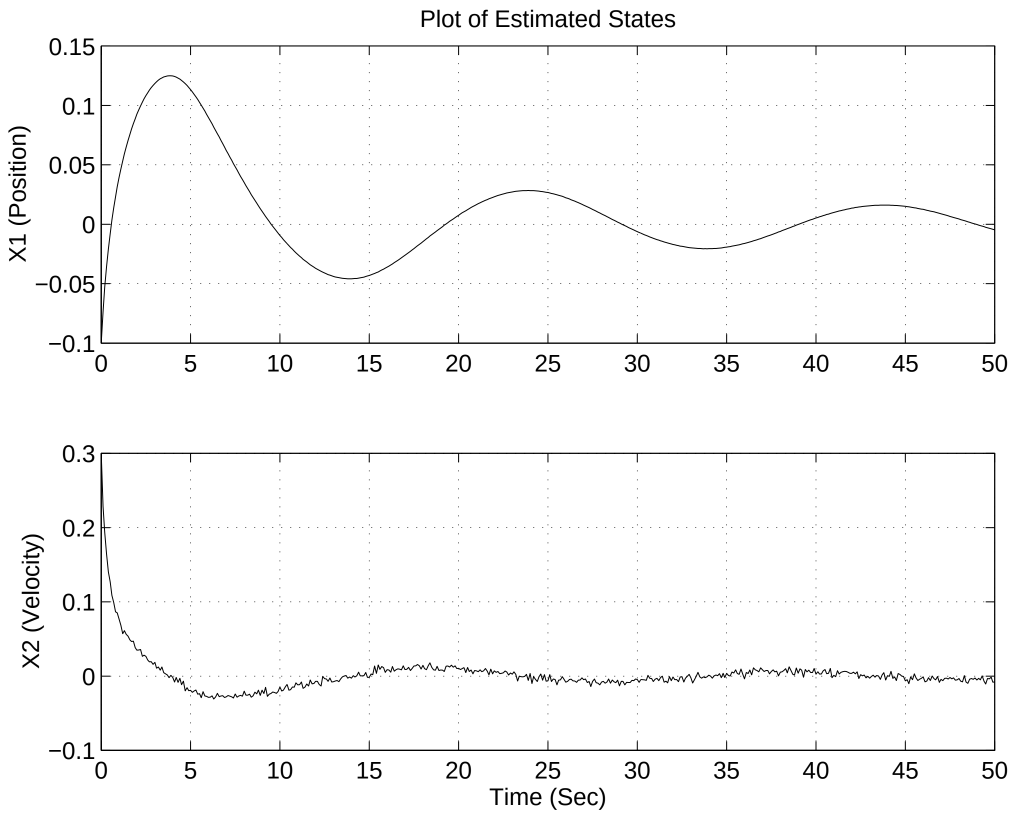
Figure 2 Estimated States

Figure 3 Actual and Determined Model Error Histories

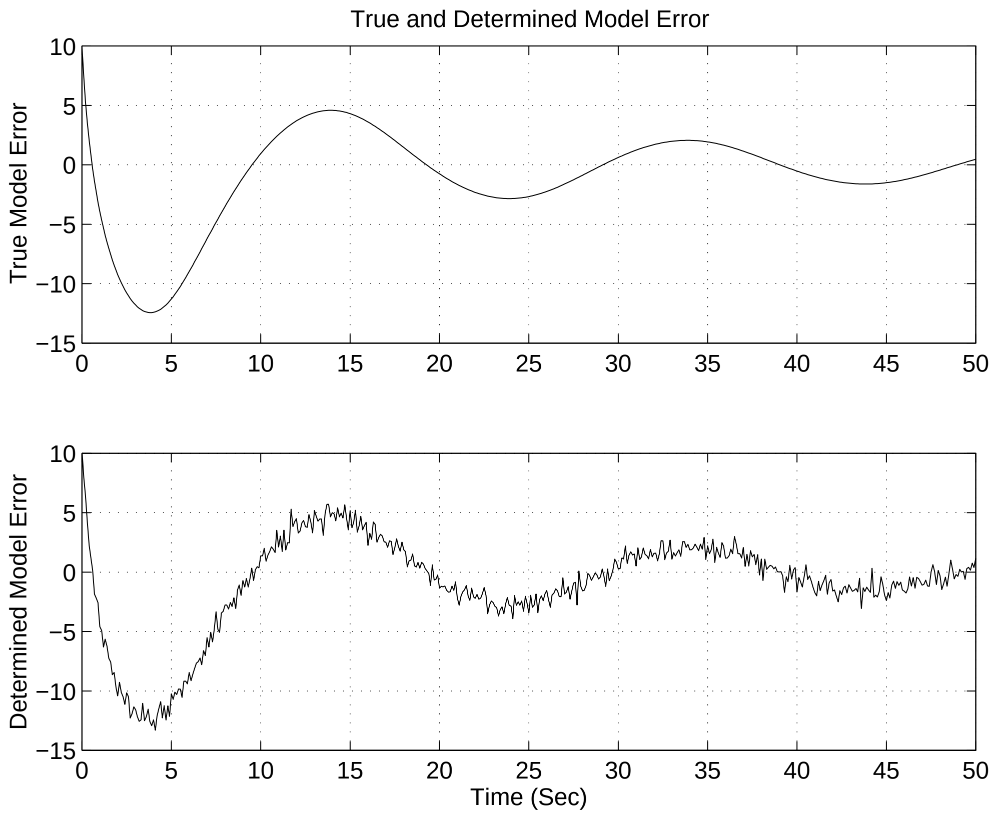
Figure 4 Sensitivity to Initial Condition Errors



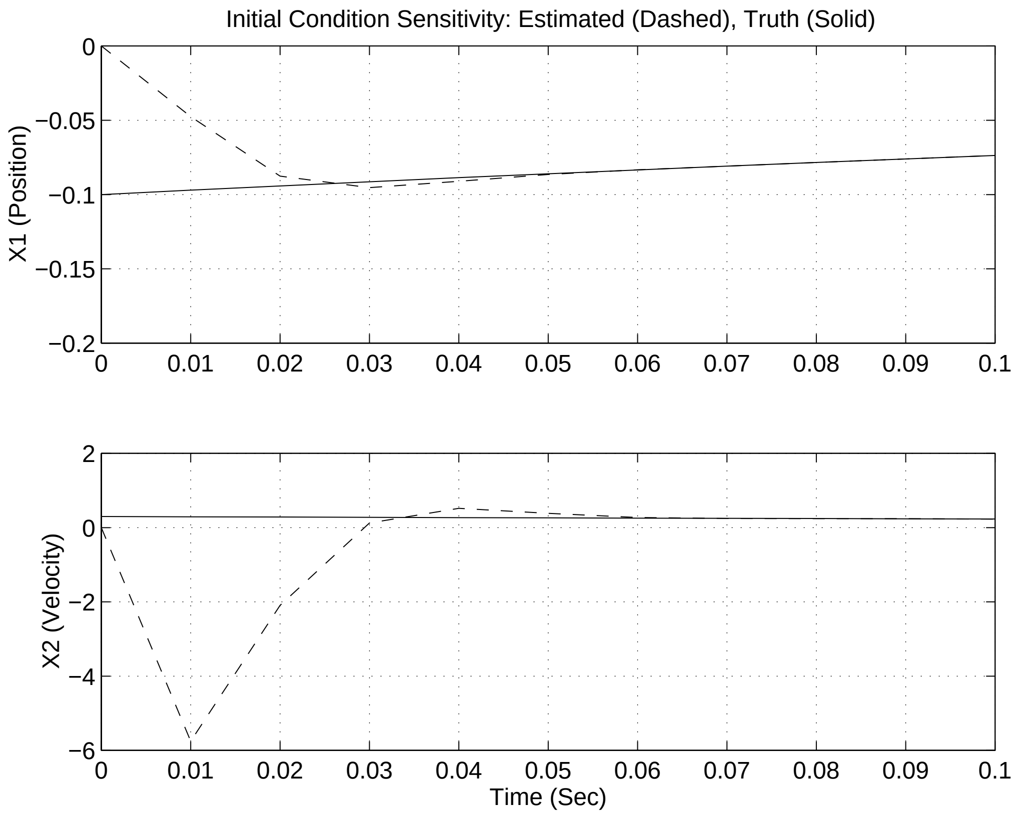
Plot of True States
X1 (Position)
0.15
0.1
0.05
0
−0.05
−0.1
0 5 10 15 20 25 30 35 40 45 50
X2 (Velocity)
0.3
0.2
0.1
0
−0.1
0 5 10 15 20 25 30 35 40 45 50
Time (Sec)



Plot of Estimated States
X1 (Position)
0.15
0.1
0.05
0
−0.05
−0.1
0 5 10 15 20 25 30 35 40 45 50
X2 (Velocity)
0.3
0.2
0.1
0
−0.1
0 5 10 15 20 25 30 35 40 45 50
Time (Sec)



True and Determined Model Error
True Model Error
Determined Model Error
Time (Sec)
10
5
0
−5
−10
−15
0
5
10
15
20
25
30
35
40
45
50


Initial Condition Sensitivity: Estimated (Dashed), Truth (Solid)
X1 (Position)
0
−0.05
−0.1
−0.15
−0.2
0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1
X2 (Velocity)
2
0
−2
−4
−6
0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1
Time (Sec)