Kalman Filter

Zhenrong
April 11, 2021

1 Linear Distrete-time Dynamic System

Consider the following linear discrete-time system

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1}, \tag{1}$$

where u_k is a known input and w_k is Gaussian zero-mean white noise with covariance Q_k. Suppose we have obtained the mean and covariance of the estimated state $\hat{x}_{k-n}$, how would $\hat{x}_k$ change over time as the k increase?

1.1 Propagation of States and Covariances

If we take the expected value of both sides of equation (1) we obtain

$$\begin{aligned} \bar{x}_k &= E\left(x_k\right) \\ &= F_{k-1}\bar{x}_{k-1} + G_{k-1}u_{k-1}. \end{aligned} \tag{2}$$

Using the definition of covariance we obtain

$$\begin{aligned} \left(x_k - \bar{x}_k\right)\left(\cdots\right)^T &= \left(F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} - \bar{x}_k\right)\left(\cdots\right)^T \\ &= \left[F_{k-1}\left(x_{k-1} - \bar{x}_{k-1}\right) + w_{k-1}\right]\left[\cdots\right]^T \\ &= F_{k-1}\left(x_{k-1} - \bar{x}_{k-1}\right)\left(x_{k-1} - \bar{x}_{k-1}\right)^T F_{k-1}^T + \\ & w_{k-1}w_{k-1}^T + F_{k-1}\left(x_{k-1} - \bar{x}_{k-1}\right)w_{k-1}^T + w_{k-1}\left(x_{k-1} - \bar{x}_{k-1}\right)^T F_{k-1}^T. \end{aligned} \tag{3}$$

Since $\left(x_{k-1} - \bar{x}_{k-1}\right)$ is uncorrelated with w_{k-1}, we obtain

$$\begin{aligned} P_k &= E\left[\left(x_k - \bar{x}_k\right)\left(\cdots\right)^T\right] \mid \\ &= F_{k-1}P_{k-1}F_{k-1}^T + Q_{k-1}. \end{aligned} \tag{4}$$

2 The Discrete-time Kalman Filter

2.1 Introduction

Suppose we have a linear discrete-time system given as follows

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1}, \\ y_k &= H_k x_k + v_k. \end{aligned} \tag{5}$$

The noise processes $\{w_k\}$ and $\{v_k\}$ are white, zero-mean, uncorrelated, and have known covariance matrices Q_k and R_k.

Our goal is to estimate the state x_k based on our knowledge of the system dynamics and the availability of the noisy measurements $\{y_k\}$. If we have all of the measurements up to and including time k available for use in our estimate of x_k, then we can form an a posteriori estimate, which we denote as $\hat{x}_k^+$. The "+" superscript denotes that the estimate is a posteriori. One way to form the a posteriori state estimate is to compute the expected value of x_k conditioned on all of the measurements up to and including time k

$$\hat{x}_k^+ = E\left[x_k \mid y_1, y_2, \cdots, y_k\right] = \text{ a posteriori estimate .} \tag{6}$$

If we have all of the measurements before (but not including) time k available for use in our estimate of x_k, then we can form an a priori estimate, which we denote as $\hat{x}_k^-$. The "-" superscript denotes that the estimate is a priori. One way to form the a priori state estimate is to compute the expected value of x_k conditioned on all of the measurements before (but not including) time k

$$\hat{x}_k^- = E\left[x_k \mid y_1, y_2, \cdots, y_{k-1}\right] = \text{ a priori estimate .} \tag{7}$$

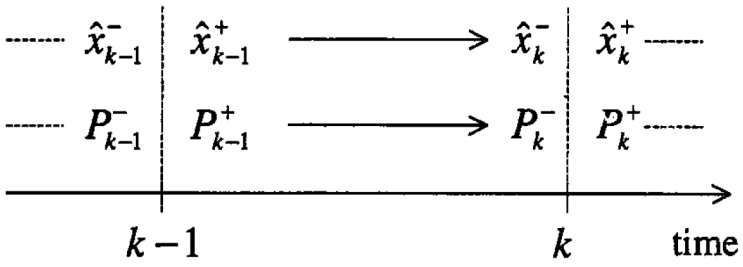

Figure 1: Timeline showing a priori and a posteriori state estimates and estimation error covariances.

The $\hat{x}_k^-$ and $\hat{x}_k^+$ are both estimates of x_k. However, $\hat{x}_k^-$ is our estimate of x_k before the measurement y_k is taken into account, and $\hat{x}_k^+$ is our estimate of x_k after the measurement y_k is taken into account. We naturally expect $\hat{x}_k^+$ to be a better estimate than $\hat{x}_k^-$, because we use more information to compute $\hat{x}_k^+$.

2.2 The Derivation of The Discrete-time Kalman Filter

2.2.1 In Least Square Sense

In least square estimation, a signal model is assumed and the probabilistic assumptions about the data is not required. The optimality criterion is to minimize the sum of the

variance of the estimation errors at time k.

A question aroused if the measurement was obtained sequentially and we want to update our estimate of x with each new measurement, we need to augment the H matrix and completely recompute the estimate $\hat{x}$. If the number of measurements becomes large, then the computational effort could become prohibitive. To address the above problem, the recursive least squares estimator was proposed. The linear recursive least squares estimation is given as follows:

1. Initialize the estimator as follows:

$$\begin{aligned} \hat{x}_0 &= E(x), \\ P_0 &= E\left[(x-\hat{x}_0)(x-\hat{x}_0)^T\right]. \end{aligned} \tag{8}$$

2. For $k = 1, 2, \cdots$, perform the following.

(a) Obtain the measurement y_k, assuming that y_k is given by the equation

$$y_k = H_k x + v_k, \tag{9}$$

where v_k is a zero-mean random vector with covariance R_k.

(b) Update the estimate of x and the estimation-error covariance P as follows:

$$\begin{aligned} K_k &= P_{k-1}H_k^T\left(H_k P_{k-1} H_k^T + R_k\right)^{-1}, \\ \hat{x}_k &= \hat{x}_{k-1} + K_k\left(y_k - H_k\hat{x}_{k-1}\right), \\ P_k &= \left(I - K_k H_k\right)P_{k-1}\left(I - K_k H_k\right)^T + K_k R_k K_k^T, \end{aligned} \tag{10}$$

where $\hat{x}_{k-1}$ and P_{k-1} are the estimate and its covariance before the measurement y_k is processed, and $\hat{x}_k$ and P_k are the estimate and its covariance after the measurement y_k is processed.

Some parts of the discrete-time Kalman filter can be directly derived from the sequential least square. In discrete-time Kalman filter, $\hat{x}_k^-$ and P_k^- are the estimate and its covariance before the measurement y_k is processed, and $\hat{x}_k^+$ and P_k^+ are the estimate and its covariance after the measurement y_k is processed. In section 2.1, we described the propagation of states and covariances and therefore the $\hat{x}_k^-$ and P_k^- can be calculated by $\hat{x}_{k-1}^+$ and P_{k-1}^+. Finally, the discrete-time Kalman filter can be given as

$$\begin{aligned} P_k^- &= F_{k-1}P_{k-1}^+F_{k-1}^T + Q_{k-1}, \\ K_k &= P_k^+ H_k^T R_k^{-1}, \\ \hat{x}_k^- &= F_{k-1}\hat{x}_{k-1}^+ + G_{k-1}u_{k-1} = \text{ a priori state estimate}, \\ \hat{x}_k^+ &= \hat{x}_k^- + K_k\left(y_k - H_k\hat{x}_k^-\right) = \text{a posteriori state estimate}, \\ P_k^+ &= \left(I - K_k H_k\right)P_k^-. \end{aligned} \tag{11}$$

2.2.2 In Bayesian Sense

The goal of a Bayesian estimator is to approximate the conditional pdf of x_k based on measurements $y_1, y_2, \cdots, y_k$, which is denoted by $p(x_k \mid Y_k)$. The first measurement is obtained at $k = 1$, so the initial condition of the estimator is the pdf of x_0, which can be written as $p(x_0)$.

Our goal is to find a recursive way to compute the conditional pdf $p(x_k \mid Y_k)$. Before we find this conditional pdf, we will find the conditional pdf $p(x_k \mid Y_{k-1})$. This is the pdf of x_k given all measurements prior to time k. We can use rewrite this pdf as

$$\begin{aligned} p(x_k \mid Y_{k-1}) &= \int p[(x_k, x_{k-1}) \mid Y_{k-1}]\, dx_{k-1} \\ &= \int p[x_k \mid (x_{k-1}, Y_{k-1})]\, p(x_{k-1} \mid Y_{k-1})\, dx_{k-1}. \end{aligned} \tag{12}$$

Noted that from our system description in Equation (1) that x_k is entirely determined by x_{k-1} and w_{k-1}. Therefore $p[x_k \mid (x_{k-1}, Y_{k-1})] = p(x_k \mid x_{k-1})$ and we see that

$$p(x_k \mid Y_{k-1}) = \int p(x_k \mid x_{k-1})\, p(x_{k-1} \mid Y_{k-1})\, dx_{k-1}. \tag{13}$$

The first pdf on the right side of the above equation is available. The pdf $p(x_k \mid x_{k-1})$ is simply the pdf of the state at time k given a specific state at time $(k-1)$. The second pdf on the right side of the above equation is a posteriori conditional pdf of x_k, which can be given as

$$p(x_k \mid Y_k) = \frac{p(y_k \mid x_k)\, p(x_k \mid Y_{k-1})}{p(y_k \mid Y_{k-1})}. \tag{14}$$

We rewrite the denominator of the above function as

$$\begin{aligned} p(y_k \mid Y_{k-1}) &= \int p[(y_k, x_k) \mid Y_{k-1}]\, dx_k \\ &= \int p[y_k \mid (x_k, Y_{k-1})]\, p(x_k \mid Y_{k-1})\, dx_k \\ &= \int p(y_k \mid x_k)\, p(x_k \mid Y_{k-1})\, dx_k. \end{aligned} \tag{15}$$

Consider what we have mentioned above, the discrete-time Kalman filter can be given as

1. The system and measurement equations are given as follows:

$$\begin{aligned} x_{k+1} &= f_k(x_k, w_k), \\ y_k &= h_k(x_k, v_k), \end{aligned} \tag{16}$$

where $\{w_k\}$ and $\{v_k\}$ are independent white noise processes with known pdf's.

2. Assuming that the pdf of the initial state $p(x_0)$ is known, initialize the estimator as follow

$$p(x_0 \mid Y_0) = p(x_0). \tag{17}$$

3. For $k = 1, 2, \cdots$,

(a) The prediction of a priori pdf is obtained from equation (11)

$$p(x_k \mid Y_{k-1}) = \int p(x_k \mid x_{k-1})\, p(x_{k-1} \mid Y_{k-1})\, dx_{k-1}. \tag{18}$$

(b) The update of a posteriori pdf is obtained from equation (12) and (13)

$$p(x_k \mid Y_k) = \frac{p(y_k \mid x_k)\, p(x_k \mid Y_{k-1})}{\int p(y_k \mid x_k)\, p(x_k \mid Y_{k-1})\, dx_k}. \tag{19}$$

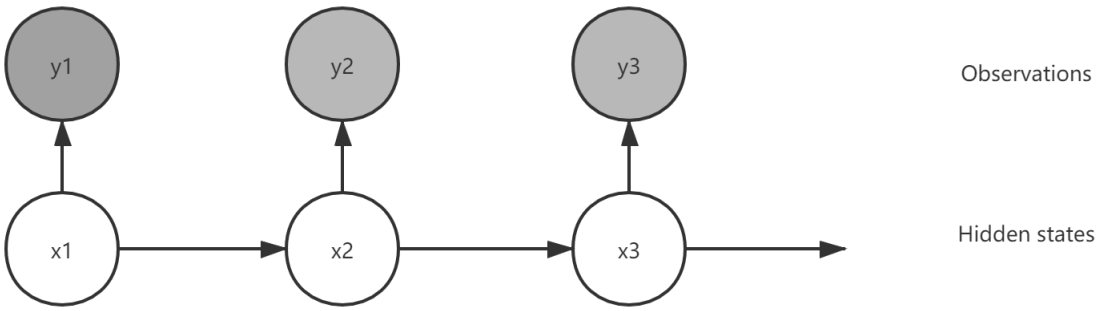

Figure 2: Illustration for a discrete-time system

Analytical solutions to these equations are available only for a few special cases. In particular, if $f(\cdot)$ and $h(\cdot)$ are linear, and $x_0, \{w_k\}$, and $\{v_k\}$ are additive, independent, and Gaussian, then the solution is the Kalman filter discussed in section 2.2.1.

2.3 Properties

- If $\{w_k\}$ and $\{v_k\}$ are Gaussian, zero-mean, uncorrelated, and white, then the Kalman filter is the optimal solution to the above problem.
- If $\{w_k\}$ and $\{v_k\}$ are zero-mean, uncorrelated, and white, then the Kalman filter is the best linear solution to the above problem.

3 Applications

3.1 Square Root Filtering

In the implementation of Kalman filter, due to the numerical problems in computer, sometimes the P_k matrices became indefinite or nonsymmetric. Square root filtering is a way to mathematically increase the precision of the Kalman filter when the computer precision is not available. By introducing the Cholesky decomposition, the conditional number of S_k, which is the square root of the P_k, is a half of which of P_k.

The results in a square root measurement-update algorithm that can be summarized as follows.

1. After the a priori covariance square root S_k^- and the a priori state estimate $\hat{x}_k^-$ have been computed, initialize

$$\begin{aligned} \hat{x}_{0k}^+ &= \hat{x}_k^-, \\ S_{0k}^+ &= S_k^-. \end{aligned} \tag{20}$$

2. For $i = 1, \cdots, r$ (where r is the number of measurements), perform the following.

(a) Define H_{ik} as the i th row of H_k, y_{ik} as the i th element of y_k, and R_{ik} as the variance of the i th measurement (assuming that R_k is diagonal).

(b) Perform the following to find the square root of the covariance after the ith measurement has been processed:

$$\begin{aligned} \phi_i &= S_{i-1,k}^{+T} H_{ik}^T, \\ a_\imath &= \frac{1}{\phi_i^T \phi_i + R_{ik}}, \\ \gamma_i &= \frac{1}{1 \pm \sqrt{a_\imath R_{ik}}}, \\ S_{ik}^+ &= S_{i-1,k}^+ \left(I - a_i \gamma_i \phi_i \phi_i^T \right). \end{aligned} \tag{21}$$

(c) Compute the Kalman gain for the i th measurement as

$$K_{ik} = a_\imath S_{ik}^+ \phi_i. \tag{22}$$

(d) Compute the state estimate update due to the i th measurement as

$$\hat{x}_{ik}^+ = \hat{x}_{i-1,k}^+ + K_{ik} \left(y_{ik} - H_{ik} \hat{x}_{i-1,k}^+ \right). \tag{23}$$

3. Set the a posteriori covariance square root and the a posteriori state estimate as

$$\begin{aligned} S_k^+ &= S_{rk}^+, \\ \hat{x}_k^+ &= \hat{x}_{rk}^+. \end{aligned} \tag{24}$$

Although square root filtering improves the numerical characteristics of the Kalman filter, it also increases computational requirements.

3.2 Optimal Smoothing (Fixed-point Smoothing)

In filtering, our objective is to obtain a posteriori estimates of x_k at time k. That is

$$\hat{x}_k = E\left(x_k \mid y_1, \cdots, y_k\right). \tag{25}$$

However, the objective in fixed-point smoothing is to obtain a posteriori state estimates of x_j at times $j+1, j+2, \cdots, k, k+1, \cdots$. We will use the notation $\hat{x}_{j,k}$ to refer to the estimate of x_j that is obtained by using all of the measurements up to and including time $(k-1)$. That is, $\hat{x}_{j,k}$ can be thought of as the a posteriori estimate of x_j at time k :

$$\hat{x}_{j,k} = E\left(x_j \mid y_1, \cdots, y_{k-1}\right), \quad k \geq j. \tag{26}$$

The trick of the derivation of the optimal smoothing is to introduce a auxiliary state $x_j^{'}$ that equal to x_j at time j and remain unchanged until time k, i.e. $x_k^{'} = x_j^{'}$.

The results in a fixed-point smoothing can be summarized as follows.

1. Run the standard Kalman filter up until time j, at which point we have $\hat{x}_{\mathrm{j}}^{-}$ and P_j^{-}.

2. Initialize the filter as follows:

$$\begin{aligned} \Sigma_j &= P_j, \\ \Pi_j &= P_j, \\ \hat{x}_{j,j} &= \hat{x}_j. \end{aligned} \tag{27}$$

3. For $k = j, j+1, \cdots$, perform the following:

$$\begin{aligned} L_k &= F_k P_k H_k^T \left(H_k P_k H_k^T + R_k\right)^{-1}, \\ \lambda_k &= \Sigma_k H_k^T \left(H_k P_k H_k^T + R_k\right)^{-1}, \\ \hat{x}_{j,k+1} &= \hat{x}_{j,k} + \lambda_k \left(y_k - H_k \hat{x}_k^{-}\right), \\ \hat{x}_{k+1}^{-} &= F_k \hat{x}_k^{-} + L_k \left(y_k - H_k \hat{x}_k^{-}\right), \\ P_{k+1} &= F_k P_k \left(F_k - L_k H_k\right)^T + Q_k, \\ \Pi_{k+1} &= \Pi_k - \Sigma_k H_k^T \lambda_k^T, \\ \Sigma_{k+1} &= \Sigma_k \left(F_k - L_k H_k\right)^T, \end{aligned} \tag{28}$$

where P_k is the a priori covariance of the standard Kalman filter estimate, Π_k is the covariance of the smoothed estimate of x_j at time k, and Σ_k is the cross covariance between the two.

3.2.1 Estimation Improvement due to Smoothing

The estimate $\hat{x}_j^-$ is the standard a priori Kalman filter estimate of x_j, and the estimate $\hat{x}_{j,k+1}$ is the smoothed estimate after measurements up to and including time k have been processed. How much more accurate can we expect our estimate to be with the use of these additional $(k+1-j)$ measurements? The improvement in estimation accuracy due to smoothing is equal to the standard estimation covariance P_j minus the smoothed estimation covariance Π_{k+1}. We can to write this improvement as

$$\begin{aligned} P_j - \Pi_{k+1} &= \Pi_j - \left(\Pi_j - \sum_{i=j}^{k} \Sigma_i H_2^T \lambda_i^T \right) \\ &= \sum_{i=j}^{k} \Sigma_i H_i^T \lambda_i^T \\ &= P \left[\sum_{t=j}^{k} \left(\tilde{F}^T \right)^{i-j} H^T \left(HPH^T + R \right)^{-1} H \tilde{F}^{i-j} \right] P. \end{aligned} \tag{29}$$

The quantity on the right side of this equation is positive definite, which shows that the smoothed estimate of x_j is always better than the standard Kalman filter estimate. Furthermore, the quantity on the right side is a sum of positive definite matrices, which shows that the larger the value of k (i.e., the more measurements that we use to obtain our smoothed estimate), the greater the improvement in the estimation accuracy. Also note from the above that the quantity $\left(HPH^T + R \right)$ inside the summation is inverted. This shows that as R increases, the quantity on the right side decreases. In the limit we see from equation 29 that

$$\lim_{R \to \infty} \left(P_j - \Pi_{k+1} \right) = 0. \tag{30}$$

This illustrates the general principle that the larger the measurement noise, the smaller the improvement in estimation accuracy that we can obtain by smoothing.

3.3 The Unscented Kalman Filter

In section 1.1, we have discussed the propagation of the states' mean and covariance in a linear dynamic system. However, in nonlinear dynamic system, this propagation process can be very complicated. Consider the following nonlinear discrete-time system

$$x_k = f(x_{k-1}) + w_{k-1}, \tag{31}$$

where w_k is Gaussian zero-mean white noise with covariance Q_k. If we take the expected value of both sided of the above equation we obtain

$$\begin{aligned}\bar{x}_k &= E\left(f(x_{k-1})\right) \\ &= \int f(x_{k-1}) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{k-1}-\mu}{\sigma}\right)^2} dx_{k-1}.\end{aligned} \tag{32}$$

For many cases, the above integral can not be compute analytically and obtaining its numerical results can be time consuming. The key idea of unscented Kalman filter is using Monte Carlo method to estimate the mean and covariance of the states, where the sample points are chosen on the basis of a specific algorithm.

3.4 Particle Filter

The particle filter was invented to numerically implement the Bayesian estimator in section 2.2.2. The results in a particle Filter can be summarized as follows.

1. The system and measurement equations are given as follows

$$\begin{aligned}x_{k+1} &= f_k\left(x_k, w_k\right), \\ y_k &= h_k\left(x_k, v_k\right),\end{aligned} \tag{33}$$

where $\{w_k\}$ and $\{v_k\}$ are independent white noise processes with known pdf's.

2. Assuming that the pdf of the initial state $p\left(x_0\right)$ is known, randomly generate N initial particles on the basis of the pdf $p\left(x_0\right)$. These particles are denoted $x_{0,2}^{+}(i = 1, \cdots, N)$. The parameter N is chosen by the user as a trade-off between computational effort and estimation accuracy.

3. For $k = 1, 2, \cdots$, do the following.

(a) Perform the time propagation step to obtain a priori particles $x_{k,i}^{-}$ using the known process equation and the known pdf of the process noise

$$x_{k,i}^{-} = f_{k-1}\left(x_{k-1,i}^{+}, w_{k-1}^{2}\right) \quad (i = 1, \cdots, N), \tag{34}$$

where each w_{k-1}^{i} noise vector is randomly generated on the basis of the known pdf of w_{k-1}.

(b) Compute the relative likelihood q_i of each particle $x_{k,i}^{-}$ conditioned on the measurement y_k. This is done by evaluating the pdf $p\left(y_k \mid x_{k,i}^{-}\right)$ on the basis of the nonlinear measurement equation and the pdf of the measurement noise.

(c) Scale the relative likelihoods obtained in the previous step as follows

$$q_i = \frac{q_i}{\sum_{j=1}^{N} q_j}. \tag{35}$$

Now the sum of all the likelihoods is equal to one.

(d) Generate a set of a posteriori particles $x_{k,i}^+$ on the basis of the relative likelihoods q_i. This is the resampling step and it is shown that the ensemble pdf of the new particles $x_{k,i}^+$ tends to the pdf $p(x_k \mid y_k)$ as the number of samples N approaches ∞.

(e) Now that we have a set of particles $x_{k,i}^+$ that are distributed according to the pdf $p(x_k \mid y_k)$, we can compute any desired statistical measure of this pdf.

3.5 Summary

In a system that is nonlinear, the Kalman filter can be used for state estimation, but the particle filter may give better results at the price of additional computational effort. In a system that has non-Gaussian noise, the Kalman filter is the optimal linear filter, but again the particle filter may perform better. The unscented Kalman filter provides a balance between the low computational effort of the Kalman filter and the high performance of the particle filter.

The particle filter has some similarities with the unscented Kalman filter in that it transforms a set of points via known nonlinear equations and combines the results to estimate the mean and covariance of the state. However, in the particle filter the points are chosen randomly, whereas in the unscented Kalman filter the points are chosen on the basis of a specific algorithm. Another difference between the two filters is that the estimation error in a unscented Kalman filter does not converge to zero, but the estimation error in a particle filter does converge to zero as the number of particles (and hence the computational effort) approaches infinity.

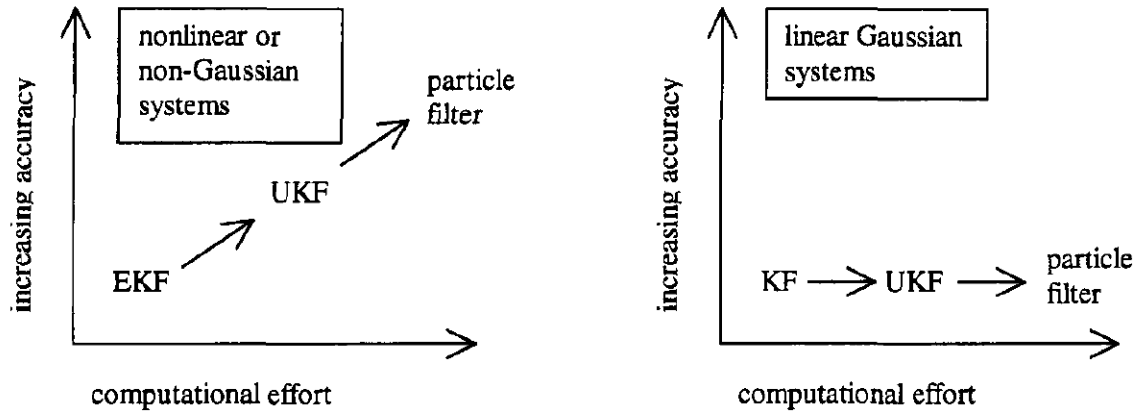

Figure 3: State estimation trade-off