Extended Kalman Filter and Particle Filter*

(COMS 4770/5770 Notes)

Yan-Bin Jia

May 25, 2026

The Kalman filter, given its linear nature, cannot be immediately used to solve most estimation problems because they are described by nonlinear systems. For such systems, the extended Kalman filter (EKF) is the most widely applied state estimation technique. Unfortunately, when the nonlinearities are high, the EKF often generates unreliable estimates due to the linearization used to propagate the mean and covariance of the state estimate. In such a situation, the particle filter is a statistical, brute-force approach to estimation that often works quite effectively. In this lecture, we will start with the linearized Kalman filter, move on to the hybrid EKF — the most common one for continuous-time dynamics and discrete-time measurements, and end with the particle filter. For other variations of the EKF and nonlinear filtering techniques (e.g., the unscented Kalman filter), we refer to Chapters 13 and 14 in [1].

1 The Continuous-Time Linearized Kalman Filter

We begin with how to linearize a nonlinear system and then apply a standard Kalman filter to estimate the deviation from a nominal state value (of the nonlinear system). Consider the following nonlinear system:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}, t), \\ \boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{v}, t), \\ \boldsymbol{w} &\sim (0, Q), \\ \boldsymbol{v} &\sim (0, R), \end{aligned} \tag{1}$$

where $\boldsymbol{u}$ is the control, $\boldsymbol{w}$ the process noise, $\boldsymbol{v}$ the measurement noise, and Q and R their respective covariance matrices.

We apply the Taylor expansion to the system (1) around nominal values of control $\boldsymbol{u}_0$, state $\boldsymbol{x}_0$, output $\boldsymbol{y}_0$, and noise $\boldsymbol{w}_0$ and $\boldsymbol{v}_0$, which are based on estimates or prior guesses. For example, if the system describes the dynamics of an airplane, then the nominal control, state, and output could represent the planned flight trajectory. The actual flight trajectory will differ from this nominal trajectory due to air disturbance and unmodeled factors, but should be close to it and within the range of approximation by the Taylor series. Without loss of generality, we assume perfectly known

*The material is adapted from Chapters 13.1–13.2 and 15 in Dan Simon's book *Optimal State Estimation* [1].

control $\boldsymbol{u}(t)$. The nominal system trajectory[1] is defined as

$$\begin{aligned} \dot{\boldsymbol{x}}_0 &= \boldsymbol{f}(\boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{w}_0, t), \\ \boldsymbol{y}_0 &= \boldsymbol{h}(\boldsymbol{x}_0, \boldsymbol{v}_0, t). \end{aligned} \tag{2}$$

We introduce

$$\begin{aligned} \Delta\boldsymbol{w} &= \boldsymbol{w} - \boldsymbol{w}_0, \\ \Delta\boldsymbol{v} &= \boldsymbol{v} - \boldsymbol{v}_0, \\ \Delta\boldsymbol{u} &= \boldsymbol{u} - \boldsymbol{u}_0 = \boldsymbol{0}, \end{aligned}$$

where the last equation is implied by the assumption.

Taylor expansion of $\boldsymbol{f}$ and $\boldsymbol{h}$ at $\xi_0 = (\boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{w}_0, t)$ leads to rewriting of the system dynamics and output:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{w}_0, t) + A\Delta\boldsymbol{x} + L\Delta\boldsymbol{w}, \\ \boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{x}_0, \boldsymbol{v}_0, t) + H\Delta\boldsymbol{x} + M\Delta\boldsymbol{v}, \end{aligned} \tag{3}$$

where

$$\begin{aligned} A &= \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right|_{(\boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{w}_0, t)}, \\ L &= \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}}\right|_{(\boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{w}_0, t)}, \\ H &= \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{(\boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{w}_0, t)}, \\ M &= \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{v}}\right|_{(\boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{w}_0, t)}. \end{aligned} \tag{4}$$

In addition, we introduce two new noise terms:

$$\begin{aligned} \tilde{\boldsymbol{w}} &= L\Delta\boldsymbol{w}, \\ \tilde{\boldsymbol{v}} &= M\Delta\boldsymbol{v}. \end{aligned}$$

It is easy to see that $\Delta\boldsymbol{w}$ and $\boldsymbol{w}$ have the same covariance, and so do $\Delta\boldsymbol{v}$ and $\boldsymbol{v}$. The two noise terms $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{v}}$ have covariance given by

$$\begin{aligned} \tilde{Q} &= LQL^\top, \\ \tilde{R} &= MRM^\top. \end{aligned} \tag{5}$$

Next, we let $\hat{\boldsymbol{x}}$ be the state estimate and introduce

$$\begin{aligned} \Delta\boldsymbol{x} &= \boldsymbol{x} - \boldsymbol{x}_0, \\ \Delta\boldsymbol{y} &= \boldsymbol{y} - \boldsymbol{y}_0, \\ \Delta\hat{\boldsymbol{x}} &= \hat{\boldsymbol{x}} - \boldsymbol{x}_0, \\ \Delta\dot{\boldsymbol{x}} &= \dot{\boldsymbol{x}} - \dot{\boldsymbol{x}}_0, \\ \Delta\dot{\hat{\boldsymbol{x}}} &= \dot{\hat{\boldsymbol{x}}} - \dot{\boldsymbol{x}}_0, \end{aligned}$$

[1] We here allow nominal noise values to be non-zero for a sense of generality and presentation consistency. In fact, we can always factor out the means of $\boldsymbol{u}$ and $\boldsymbol{v}$ and rewrite $\boldsymbol{f}$ and $\boldsymbol{h}$ such that the remaining parts of the two types of noise have zero mean values. This will essentially set $\boldsymbol{w}_0(t) = \boldsymbol{0}$ and $\boldsymbol{v}_0(t) = \boldsymbol{0}$.

Subtracting the two equations in (2) from those in (3), respectively, we obtain a linearized system whose state is the estimation error of the state of the original system:

$$\begin{array}{rcl} \Delta\dot{\hat{\boldsymbol{x}}} &=& A\Delta\hat{\boldsymbol{x}} + \tilde{\boldsymbol{w}}, \\ \Delta\boldsymbol{y} &=& H\Delta\boldsymbol{x} + \tilde{\boldsymbol{v}}, \\ \tilde{\boldsymbol{w}} &\sim& \tilde{Q}, \\ \tilde{\boldsymbol{v}} &\sim& \tilde{R}. \end{array} \tag{6}$$

The Kalman filter for (6) then follows from the standard version for a linear system[2]:

$$\begin{array}{rcl} \Delta\dot{\hat{\boldsymbol{x}}} &=& A\Delta\hat{\boldsymbol{x}} + K\left(\Delta\boldsymbol{y} - H\Delta\hat{\boldsymbol{x}}\right), \\ K &=& PH^T\tilde{R}^{-1}, \\ \dot{P} &=& AP + PA^T + \tilde{Q} - PH^\top\tilde{R}^{-1}HP. \end{array} \tag{7}$$

It is initialized below:

$$\begin{array}{rcl} \Delta\hat{\boldsymbol{x}}(0) &=& \mathbf{0}, \\ P(0) &=& E\left(\Big(\Delta\boldsymbol{x}(0) - \Delta\hat{\boldsymbol{x}}(0)\Big)\Big(\Delta\boldsymbol{x}(0) - \Delta\hat{\boldsymbol{x}}(0)\Big)^\top\right). \end{array} \tag{8}$$

We just need add the state update:

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_0 + \Delta\hat{\boldsymbol{x}}. \tag{9}$$

2 The Extended Kalman Filter

The linearized Kalman filter (7)–(9) requires knowledge of a nominal state trajectory $\boldsymbol{x}_0(t)$ for evaluation of the matrices $A, H, \tilde{Q}$, and $\tilde{R}$. The extended Kalman filter (EKF) has the idea of using the filter estimate as the nominal state trajectory. In other words, we choose $\boldsymbol{x}_0(t) = \hat{\boldsymbol{x}}(t)$ so that $\Delta\hat{\boldsymbol{x}}(t) = \mathbf{0}$ and $\Delta\dot{\hat{\boldsymbol{x}}}(t) = \mathbf{0}$. The four matrices A, L, H, M in (4) are obtained from evaluating the corresponding partial derivatives at $(\hat{\boldsymbol{x}}, \boldsymbol{u}, \boldsymbol{w}, t)$. The EKF consists of the following equations:

$$\begin{array}{rcl} \hat{\boldsymbol{x}}(0) &=& E(\boldsymbol{x}(0)), \\ P(0) &=& E\left((\boldsymbol{x}(0) - \hat{\boldsymbol{x}}(0))(\boldsymbol{x}(0) - \hat{\boldsymbol{x}}(0))^\top\right), \\ \dot{\hat{\boldsymbol{x}}} &=& \boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{u}, \boldsymbol{w}_0, t) + K\left(\boldsymbol{y} - \boldsymbol{h}(\hat{\boldsymbol{x}}, \boldsymbol{v}_0, t)\right), \\ K &=& PH^T\tilde{R}^{-1}, \\ \dot{P} &=& AP + PA^T + \tilde{Q} - PH^\top\tilde{R}^{-1}HP, \end{array} \tag{10}$$

Compared with the linearized Kalman filter (7)–(9), changes happen mainly to the first equation in (7) describing $\Delta\dot{\hat{\boldsymbol{x}}}$, the derivative of the state estimation discrepancy. We replace the linearization term $\boldsymbol{x}_0 + A\Delta\hat{\boldsymbol{x}}$ with $\boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{u}, \boldsymbol{w}_0, t)$, which is more accurate. Also for accuracy, $\boldsymbol{h}$ is used in the feedback term.

The system (10) is referred to as the *continuous-time EKF*. It assumes not only continuous-time dynamics but also measurements available continuously in time. It is true that many real systems, mechanical or electromagnetic, are described by differential equations that are continuous in time. However, measurements by sensors can only be conducted at discrete time instants.

[2]Refer to the lecture notes titled "Discrete- and Continuous-Time Kalman Filters".

The common situation in practice prompts us to consider the following continuous-time system with discrete-time measurements:

$$\begin{array}{rcl} \dot{\boldsymbol{x}} & = & \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}, t), \\ \boldsymbol{y}_k & = & \boldsymbol{h}_k(\boldsymbol{x}_k, \boldsymbol{v}_k), \\ \boldsymbol{w}(t) & \sim & (0, Q), \\ \boldsymbol{v}_k & \sim & (0, R_k). \end{array} \tag{11}$$

where the process noise $\boldsymbol{w}(t)$ is continuous-time white noise with covariance Q and the measurement noise $\boldsymbol{v}_k$ at time instant t_k is white noise with covariance R_k.

Between measurements, we propagate the state estimate according to the nonlinear dynamics equation for $\dot{\boldsymbol{x}}$ above:

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{u}, \boldsymbol{w}_0, t).$$

We also propagate the covariance according to the equation for $\dot{P}$ in (10), which is expanded via a substitution of (5) as

$$\dot{P} = AP + PA^T + \tilde{Q} - PH^\top (MRM^\top)^{-1} HP.$$

During the integration between two measurements we make zero measurement, or equivalently, we make measurements with infinite covariance $R = \infty I$. The last term on the right hand side of the above equation goes to zero accordingly.

At each measurement time, we update the state estimate and the covariance from their prior values to posterior values using the discrete-time Kalman filter (see the notes titled "Discrete- and Continuous-Time Kalman Filters").

Putting the pieces together, we have derived the most commonly used version of the EKF — for the system (11) — as follows.

1. Initialize the filter:

$$\begin{array}{rcl} \hat{\boldsymbol{x}}_0^+ & = & E(\boldsymbol{x}_0), \\ P_0^+ & = & E\left((\boldsymbol{x}_0 - \boldsymbol{x}_0^+)(\boldsymbol{x}_0 - \boldsymbol{x}_0^+)^\top \right). \end{array}$$

2. Repeat the following for every step $k = 1, 2, \ldots,$

   a) Update the state estimate and its covariance from time $(k-1)^+$ to time k^- via integrating the two equations below:

$$\begin{array}{rcl} \dot{\hat{\boldsymbol{x}}} & = & \boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{u}, \boldsymbol{w}_0, t), \\ \dot{P} & = & AP + PA^\top + LQL^\top. \end{array}$$

   In the above, A and L are evaluated according to (4).

   b) Update the state estimate and its covariance from time k^- to time k^+ by incorporating the measurement $\boldsymbol{y}_k$:

$$\begin{array}{rcl} K_k & = & P_k^- H_k^\top \left(H_k P_k^- H_k^\top + M_k R_k M_k^\top \right)^{-1}, \\ \hat{\boldsymbol{x}}_k^+ & = & \hat{\boldsymbol{x}}_k^- + K_k \Big(\boldsymbol{y}_k - \boldsymbol{h}_k(\hat{\boldsymbol{x}}_k^-, \boldsymbol{v}_k, t_k) \Big), \\ P_k^+ & = & (I - K_k H_k) P_k^- (I - K_k H_k)^\top + K_k M_k R_k M_k^\top K_k^\top, \end{array}$$

   where H_k and M_k are the partial derivatives of $\boldsymbol{h}_k(\boldsymbol{x}_k, \boldsymbol{v}_k)$ with respect to $\boldsymbol{x}_k$ and $\boldsymbol{v}_k$, respectively, and are both evaluated at $\hat{\boldsymbol{x}}_k^-$.

In the above, the effects of linearization are evidenced in the occurrences of $LQL^\top$ and $M_k R_k M_k^\top$. The algorithm is referred to as the *hybrid EKF*.

EXAMPLE 1. We use the continuous-time and hybrid EKFs to estimate the motion of a body as it falls toward the earth. The state consists of the body's altitude x_1, velocity x_2, and x_3, whose inverse is the constant ballistic coefficient. The altitude can be measured using a range device. The system of equations is given as

$$\begin{aligned} \dot{x}_1 &= x_2 + w_1, \\ \dot{x}_2 &= \frac{\rho_0 x_2^2}{2x_3} e^{-x_1/k} - g + w_2, \\ \dot{x}_3 &= w_3, \\ y &= x_1 + v, \end{aligned} \tag{12}$$

where w_1, w_2, w_3 are the process noise, v the measurement noise, ρ_0 the air density at sea level, k some constant relating air density and altitude, and g the gravitational acceleration.

We obtain the partial derivatives of h and $\boldsymbol{f}$ with respect to $\boldsymbol{x} = (x_1, x_2, x_3)$:

$$\begin{aligned} A &= \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \\ &= \begin{pmatrix} 0 & 1 & 0 \\ -\dfrac{\rho_0 x_2^2}{2kx_3} e^{-x_1/k} & \dfrac{\rho_0 x_2}{x_3} e^{-x_1/k} & -\dfrac{\rho_0 x_2^2}{2x_3^2} e^{-x_1/k} \\ 0 & 0 & 0 \end{pmatrix}, \\ H &= (1, 0, 0). \end{aligned}$$

The constants are given below:

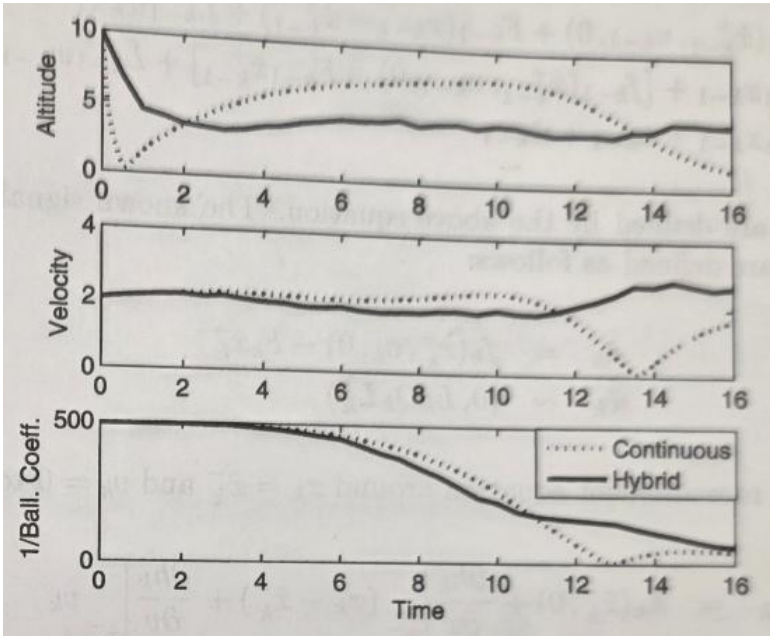
$$\begin{aligned} \rho_0 &= 0.0034 \,\text{lb-sec}^2/\text{ft}^4, \\ g &= 32.2 \,\text{ft/sec}^2, \\ k &= 22000 \,\text{ft}, \\ E(v^2(t)) &= 100 \,\text{ft}^2, \\ E(w_i^2(t)) &= 0, \qquad i = 1, 2, 3. \end{aligned}$$

Initialize the system and estimator as

$$\boldsymbol{x}_0 = \begin{pmatrix} 100,000 \\ -6,000 \\ 1/2,000 \end{pmatrix}, \qquad \hat{\boldsymbol{x}}_0^+ = \begin{pmatrix} 100,010 \\ -6,100 \\ 1/2,500 \end{pmatrix}, \qquad \text{and} \qquad P_0^+ = \begin{pmatrix} 500 & 0 & 0 \\ 0 & 20,000 & 0 \\ 0 & 0 & 1/250,000 \end{pmatrix}.$$

This gives enough information to simulate the continuous-time EKF through numerical integration. To proceed with the hybrid EKF, we suppose that range measurements are taken every 0.5 seconds. Simulation results are plotted in the next figure.[3] The continuous EKF performs in general better than the hybrid EKF largely because the former makes more measurements (a continuum in theory) than the latter. In practice, the continuous EKF has to use discretization subject to the frequency of measurements, essentially leveled with the hybrid EKF. Propagation from t_k^+ to t_{k+1}^- assumes finite covariance R, which tends to introduce more errors than the hybrid EKF.

[3] Figure 13.3 on p. 407 of [1].



3 The Particle Filter

The particle filter is a nonlinear state estimator that trades computational effort for the degree of nonlinearity. A probability-based estimator, it has proven very effective in machine learning, mapping, and learning robots [2].

At the beginning of estimation, we randomly generate a number of state vectors based on the probability distribution of the initial state (which is assumed to be known). These state vectors are called *particles*. At each time step, we propagate the particles to the next time step using the process equation. Compute the likelihood of each particle based on the measurement taken at the step. Then generate a set of new particles based on the computed relative likelihoods. Iterate for a number of steps. Output the mean and covariance of the particles at the last step.

More formally, suppose the nonlinear system and measurement equations are given as follows for the time step i:

$$\begin{aligned} \boldsymbol{x}_{i+1} &= \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{w}_i), \\ \boldsymbol{y}_i &= \boldsymbol{h}_i(\boldsymbol{x}_i, \boldsymbol{v}_i). \end{aligned}$$

where $\boldsymbol{w}_i$ and $\boldsymbol{v}_i$ are the process noise and measurement noise, respectively, with known probability density functions (pdf's). The functions $\boldsymbol{f}_i$ and $\boldsymbol{h}_i$ vary with time (i.e., the time step). The pdf $p(\boldsymbol{x}_0)$ of the initial state is assumed to be known.

The particle filtering algorithm executes the following steps:

1. Randomly generate r initial particles based on $p(\boldsymbol{x}_0)$, and denote them as $\boldsymbol{x}_{0,j}^+$, $j = 1, \ldots, r$. The parameter r is chosen as a trade-off between computational effort and estimation accuracy.

2. Repeat the following for each step $i = 1, 2, \ldots,$.

Propagation For each particle $\boldsymbol{x}_{i-1,j}^{+}$ perform the time propagation step to obtain *a priori* particles $\boldsymbol{x}_{i,j}^{-}$, based on the system dynamics:

$$\boldsymbol{x}_{i,j}^{-} = \boldsymbol{f}_{i-1}(\boldsymbol{x}_{i-1,j}^{+}, \boldsymbol{w}_{i-1}^{j}), \qquad \text{for } j = 1, \ldots, r,$$

where each noise vector $\boldsymbol{w}_{i-1}^{j}$ is randomly generated based on the known pdf of $\boldsymbol{w}_{i-1}$.

Likelihood determination Compute the relative likelihood q_j of each particle $\boldsymbol{x}_{i,j}$, conditioned on the measurement $\boldsymbol{y}_i$. This is done by evaluating the pdf $p(\boldsymbol{y}_i \mid \boldsymbol{x}_{i,j}^{-})$, since we know the pdf of the measurement noise. For instance, if an m-dimensional measurement is given as

$$\boldsymbol{y}_i = \boldsymbol{h}(\boldsymbol{x}_i) + \boldsymbol{v}_i,$$

where $\boldsymbol{v}_i \sim N(0, R)$, then the relative likelihood of a specific measurement $\boldsymbol{y}^*$ can be computed as follows:

$$\begin{aligned} q_j &= \Pr(\boldsymbol{y}_i = \boldsymbol{y}^* \mid \boldsymbol{x} = \boldsymbol{x}_{i,j}^{-}) \\ &= \Pr(\boldsymbol{v}_i = \boldsymbol{y}^* - \boldsymbol{h}(\boldsymbol{x}_{i,j}^{-})) \\ &\sim \frac{1}{(2\pi)^{m/2}|R|^{1/2}} \exp\left(\frac{-(\boldsymbol{y}^* - \boldsymbol{h}(\boldsymbol{x}_{i,j}^{-}))^{\top} R^{-1} (\boldsymbol{y}^* - \boldsymbol{h}(\boldsymbol{x}_{i,j}^{-}))}{2} \right). \end{aligned} \tag{13}$$

The $\sim$ symbol means that the probability is proportional to the right hand side. We use it to compare one relative likelihood with another because all are scaled by the same factor.

Normalization Scale the relative likelihoods (13) obtained in the previous step as follows:

$$q_j = \frac{q_j}{\sum_{k=1}^{r} q_k}.$$

After the scaling, all the likelihoods add up to one.

Resampling Generate r posterior particles $\boldsymbol{x}_{i,j}^{+}$ based on the relative likelihoods q_j.

In each iteration step, we have a set of particles $\boldsymbol{x}_{i,j}^{+}$ distributed according to the pdf $p(x_i \mid y_i)$. We can compute the mean and covariance.

EXAMPLE 2. We consider a benchmark problem in nonlinear estimation where the system is given as follows:

$$\begin{aligned} x_i &= \frac{1}{2} x_{i-1} + \frac{25 x_{i-1}}{1 + x_{i-1}^2} + 8\cos(1.2(i-1)) + w_i, \\ y_i &= \frac{1}{20} x_i^2 + v_i, \end{aligned}$$

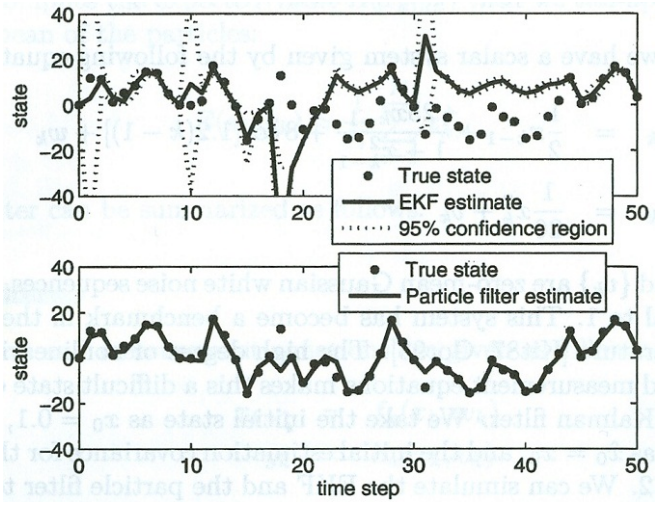
where $\{w_i\}$ and $\{v_i\}$ are Gaussian white noise sequences with zero means, and unit variances.

The high degree of nonlinearity in the process and the measurement makes the problem difficult for a Kalman filter. We take the initial state $x_0 = 0.1$ and initial estimate $\hat{x}_0 = x_0$, and the initial estimation covariance for the Kalman filter as $P_0^{+} = 2$. Simulate the extended Kalman filter (EKF) [1] and the particle filter over 50 time steps (using $r = 100$ particles).

The next figure[4] compares the EKF and particle filter estimates. Not only is the EKF estimate poor, but also the EKF thinks (based on the computed covariance) that the estimate is better than it really is. In

[4]Figure 15.3 on p. 470 of [1]

comparison, the particle filter does a better job. The estimation errors[5] for the Kalman and particle filters are 16.3 and 2.6, respectively.



References

[1] D. Simon. *Optimal State Estimations.* John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[2] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics.* The MIT Press, Boston, MA, 2005.

[5] measured as the roots of the mean squares.