2.160 IDENTIFICATION, ESTIMATION, AND LEARNING

LECTURE NOTES NO. 7

7. Extended Kalman Filter and Unscented Kalman Filter

7.1 Nonlinear Process Dynamics

In many practical problems, the process dynamics are nonlinear.

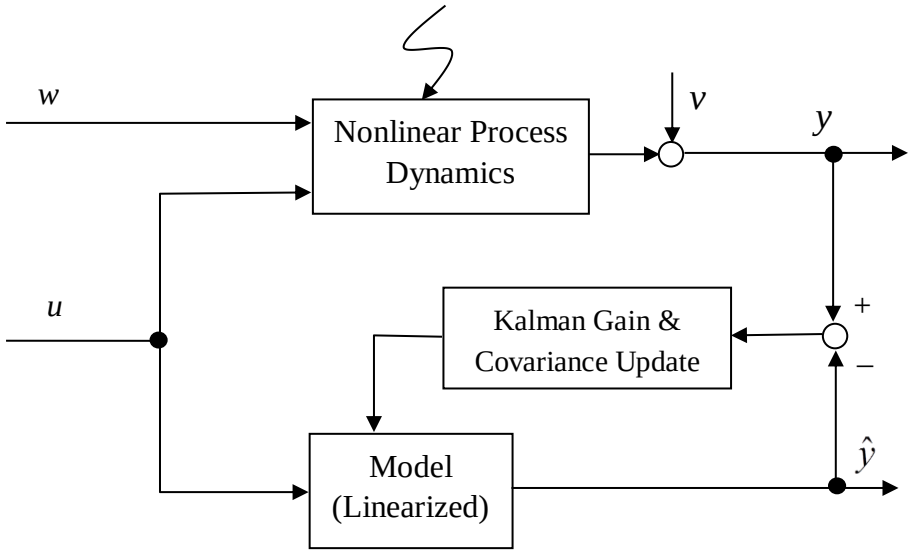


Figure 7.1 Kalman filter for nonlinear systems

If the process is nonlinear but smooth, its linearized approximation may be used for the process model.

Consider a non-linear, continuous system

$$\dot{x} = f(x,u,t) + w(t) \quad \ldots \quad n-\dim$$

$$y = h(x,t) + v(t) \quad \ldots \quad l-\dim \tag{1}$$

f (.), and *h(.)*: known but non-linear, differentiable functions

u : input (deterministic forcing term; assumed zero)

w, v : uncorrelated process and measurement noises

$$E[w(t)] = 0, \quad E[v(t)] = 0$$

$$E[w(t)w^T(s)] = \begin{cases} 0 & t \neq s \\ Q & t = s \end{cases} \qquad E[v(t)v^T(s)] = \begin{cases} 0 & t \neq s \\ Q & t = s \end{cases} \qquad E[w(t)v^T(s)] = 0$$

The original Kalman filter is not applicable to this class of nonlinear systems. This section describes the extension of Kalman Filtering to nonlinear systems.

Approaches :

7.2 Linearize the non-linear system along a trajectory to track
…Linearized Kalman Filter

7.3 Linearize it around the estimated state in real-time for covariance update + nonlinear output prediction and state transition
…Extended Kalman Filter.

7.4 Use a statistical sampling technique for covariance update
+ nonlinear output prediction and state transition
…Unscented Kalman Filter

7.2 Linearized Kalman Filter

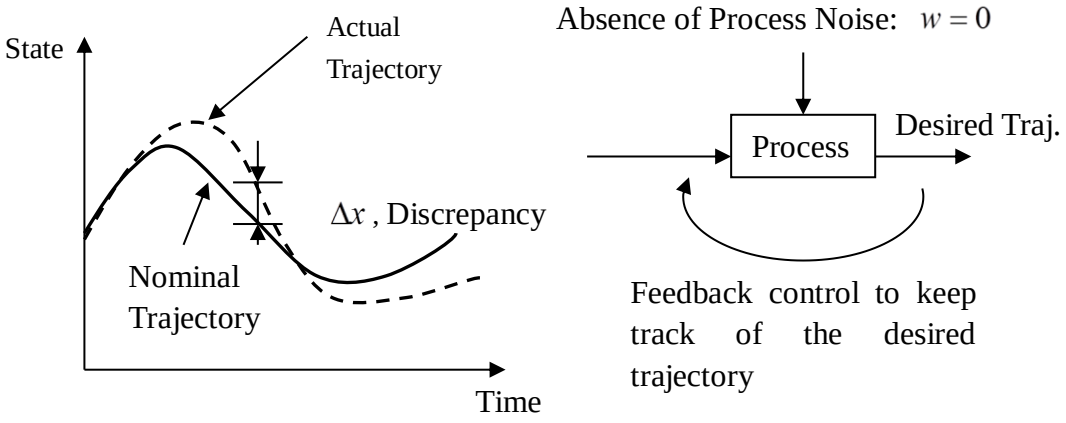


Figure 7.2 Nominal trajectory

$x^*(t) = a$ nominal trajectory in the state space satisfying the noise-less state equation:

$$\dot{x}^*(t) = f\left(x^*(t), t\right) \tag{2}$$

absence of process noise

Consider deviation, $\Delta x(t)$, from the nominal trajectory

$$x(t) = x^*(t) + \Delta x(t) \tag{3}$$

$$\dot{x}(t) = \dot{x}^*(t) + \Delta \dot{x}(t) \tag{4}$$

Taylor expansion

$$\begin{aligned} f(x,t) &= f\left(x^* + \Delta x, t\right) \\ &\cong f\left(x^*, t\right) + \frac{\partial f}{\partial x}\Big|_{x=x^*} \Delta x \end{aligned} \tag{5}$$

where

$$\left.\frac{\partial f}{\partial x}\right|_{x=x^*} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \ddots & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial f_n}{\partial x_1} & \cdots & & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_{x=x^*} \in R^{n\times n} \quad \text{Jacobian} \tag{6}$$

Combining (4) and (5)

$$\dot{x}^* + \Delta\dot{x} = \underbrace{f(x^*, t)}_{\dot{x}^* \text{ from (2)}} + \left.\frac{\partial f}{\partial x}\right|_{x^*} \Delta x + w(t)$$

$$\Delta\dot{x} = \left.\frac{\partial f}{\partial x}\right|_{x^*} \Delta x + w(t)$$

Now replacing Δx by x and $\left.\frac{\partial f}{\partial x}\right|_{x^*}$ by $F(t) \in R^{n\times n}$

$$\dot{x} = F(t)x + w(t) \tag{7}$$

Similarly, from (1),

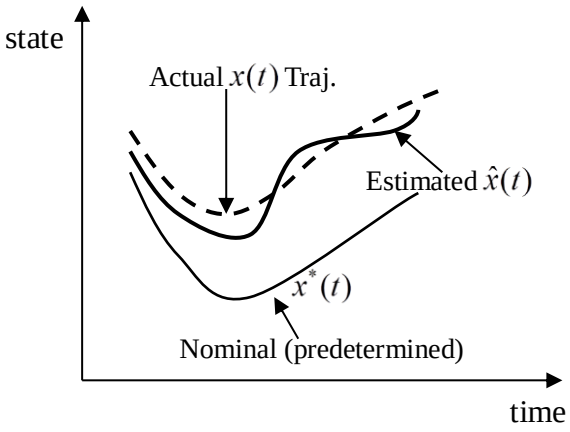
$$y = H(t)x + v(t) \qquad H(t) = \left.\frac{\partial h}{\partial x}\right|_{x^*} \in R^{l\times n} \tag{8}$$

$$y^* + \Delta y = \underbrace{h(x^*, t)}_{y^*} + \underbrace{\left.\frac{\partial h}{\partial x}\right|_{x^*}}_{H(t)} \Delta x + v$$

Replacing Δy by y

Note that the above linearized system (7), (8) with $F(t)$ and $H(t)$ are in the same form as that of the original Kalman filter: linear time-varying systems.

7.3 Extended Kalman Filter



With the measurement of the actual process, the estimated trajectory is deemed to be more accurate. Use the state $\hat{x}(t)$ estimated in real-time for linearizing the dynamics:

$$\left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}} = F(\hat{x},t)$$

$$\left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} = H(\hat{x},t) \qquad (9)$$

Figure 7.3 Actual, estimated, and nominal trajectories

Namely, matrices *F* and *H* are evaluated at $x=\hat{x}$, the estimated values of the state in real time, rather than its nominal values. Note that $F(\hat{x},t)$ and $H(\hat{x},t)$ cannot be pre-computed in off-line.

For estimating the output, however, we do not have to use the linearized model; the nonlinear output function, eq.(1), can be used:

$$\hat{y}(t) = h(\hat{x}(t),t) \qquad (10)$$

The state propagation, too, can be replaced by the original nonlinear function: $f(\hat{x}(t))$.

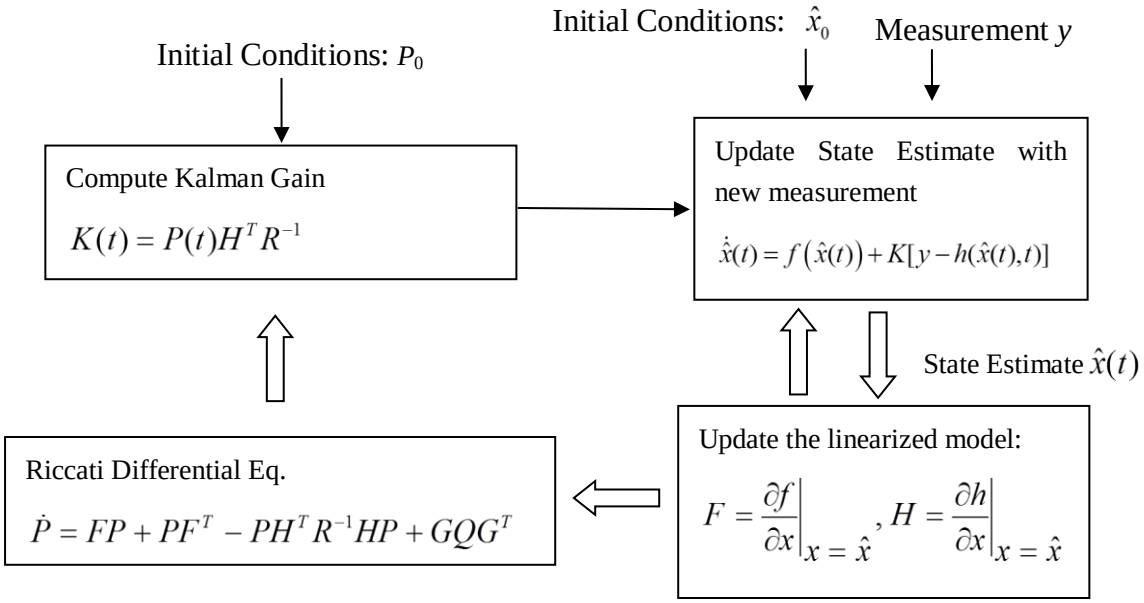


Figure 7.4 Extended Kalman Filter

The Extended Kalman Filter is a type of linearized Kalman filter with the above two

modifications, (9) and (10), for better use of the nonlinear dynamics. See Figure 7.4. A critical issue of this Extended Kalman Filter is instability. As estimated state $\hat{x}$ deviates from the true state, the linearized model becomes inaccurate, which may lead to an even larger error in state estimation. Care must be taken in implementing the extended Kalman Filter.

7.4 Unscented Transform

Although the Extended Kalman Filter works satisfactorily for many applications, it is known that it performs poorly when the system is highly nonlinear, exhibiting large deviations from the piece-wise linearized dynamics. Since the covariance update of the Extended Kalman Filter is based on the linearized model with

$$F = \left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}}, H = \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}}, \tag{11}$$

the resultant error covariance *P* does not reflect the true value. In particular, it is known that the Extended Kalman filter tends to underestimate the error covariance when the Jacobian-based linear model (11) changes rapidly, resulting in a smaller Kalman gain and insufficient state update. This often leads to divergence of the Extended Kalman filter. See Figure 7.5 below.

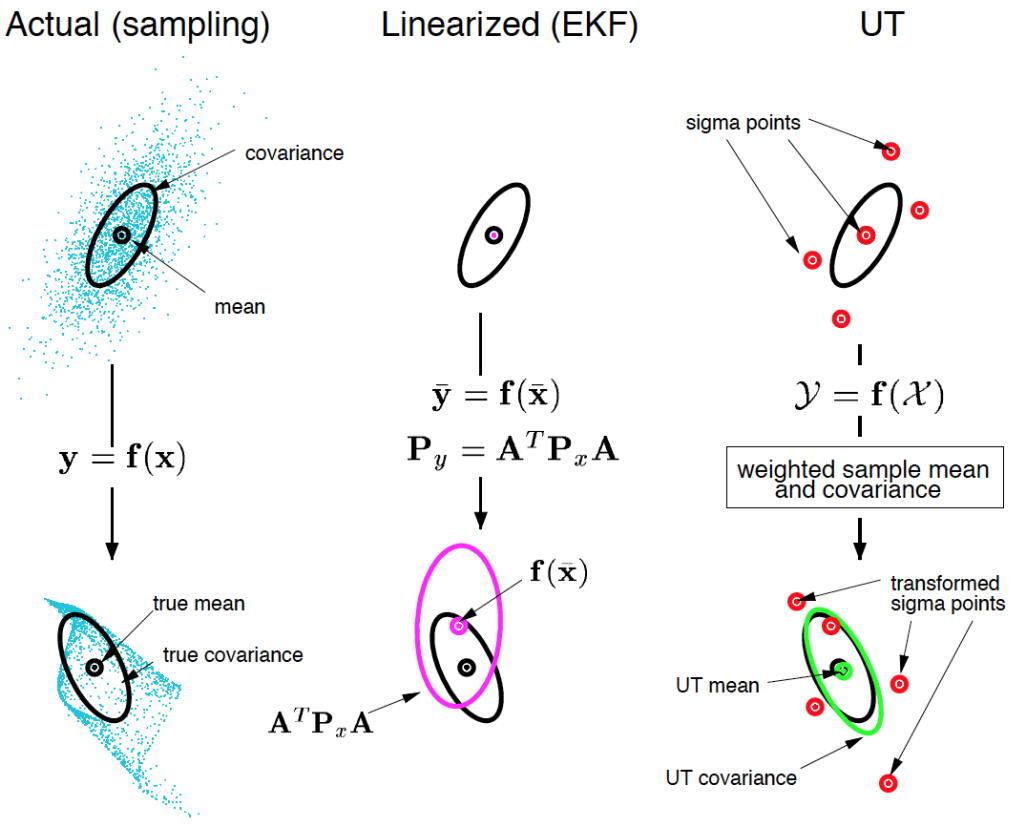


Figure 7.5 Mean and covariance simulated for a) the true nonlinear system, b) EKF using linearized state transition matrix A, and c) Unscented Transformation. From Eric A. Wan and Rudolph van der Merwe, "The Unscented Kalman Filter for Nonlinear Estimation"

This problem can be solved by using Unscented Kalman Filter by Julier and Uhlmann [1997], which estimates more accurate error covariance despite prominent nonlinearity. The key idea is to estimate the error covariance based on a special set of sample points, termed "sigma points", which propagate directly through the original nonlinear model, rather than computing it based on the covariance update and propagation formula, (4-41) and (4-45), or the Riccati equation, which need the linearized process model, (4-5) and (4-6) or (11). Sigma points are an effective sampling technique that provides accurate mean and covariance values with a minimum number of sample points. These sigma points are propagated through the nonlinear dynamic model to estimate the error covariance of $\hat{x}_{t+1|t}$. This results in a filter capturing more accurate mean and covariance despite the nonlinearity. Furthermore, this Unscented Kalman Filter does not require to analytically compute the Jacobians, which for complex systems is difficult to perform.

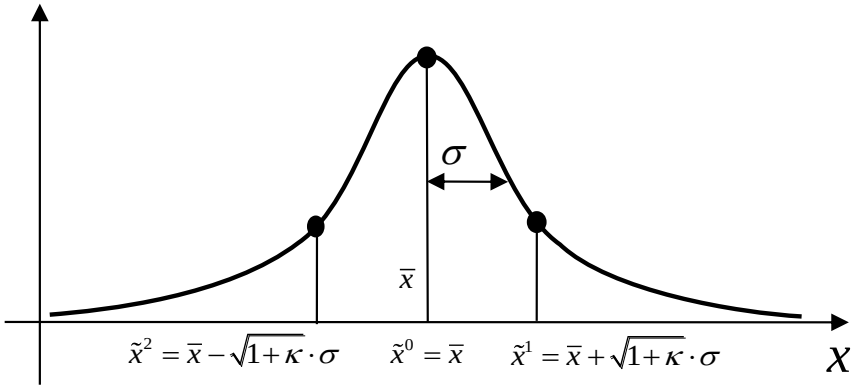


Figure 7-6 Gaussian distribution and sigma points

To begin let us consider a simple scalar case. Assume that a scalar random variable $x \in X$ has a Gaussian distribution, which is completely characterized by mean $\bar{x}$ and variance σ^2. Take three sample points, called "sigma points", as follows:

$$\begin{aligned} \tilde{x}^0 &= \bar{x} & W_0 &= \frac{\kappa}{1+\kappa} \\ \tilde{x}^1 &= \bar{x} + \sqrt{1+\kappa}\cdot\sigma & & \\ \tilde{x}^2 &= \bar{x} - \sqrt{1+\kappa}\cdot\sigma & W_1 &= W_2 = \frac{1}{2(1+\kappa)} \end{aligned} \tag{12}$$

where κ is a parameter of sigma points to be tuned, and W_i is weight of the *i-th* sigma point. See Figure 7-6. The weighted mean and variance of these samples agree with the mean and variance of the original distribution of *x*.

$$\sum_{i=0}^{2} W_i \tilde{x}^i = \bar{x} \tag{13}$$

$$\sum_{i=0}^{2} W_i (\tilde{x}^i - \bar{x})^2 = \frac{1}{2(1+\kappa)}\left((\sqrt{1+\kappa}\cdot\sigma)^2 + (-\sqrt{1+\kappa}\cdot\sigma)^2\right) = \sigma^2$$

Consider a nonlinear, differentiable function that can be represented by a Taylor series expansion, i.e. an analytic function:

$$y = f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\bar{x})}{k!}(x-\bar{x})^k \tag{14}$$

Note that $y \in Y$ is a random variable as x is an instantiation of random variable X; $x \in X$. Construct a set of sample points in y that correspond to the individual sigma points in x.

$$\tilde{y}^i = f(\tilde{x}^i), \quad i = 0,1,2 \tag{15}$$

Compute the weighted mean and variance of the transformed sigma points: $\{\tilde{y}_i \mid i = 0,1,2\}$.

$$\bar{y}_{sample} = \sum_{i=0}^{2} W_i \tilde{y}^i$$
$$\sigma^2_{sample} = \sum_{i=0}^{2} W_i (\tilde{y}^i - \bar{y}_{sample})^2 \tag{16}$$

The weighted mean of the sigma points transformed through the nonlinear analytic function approximates the true mean to the third order, and that of the covariance to the second order. [A problem in PS]. In other words, the transformed sigma points $\{\tilde{y}_i \mid i = 0,1,2\}$ approximate the true distribution of Y to.

$$\bar{y}_{sample} = E[y] + \mathrm{O}(4), \quad \sigma^2_{sample} = E[(y - E[y])^2] + \mathrm{O}(3) \tag{17}$$

For a general n-dimensional multivariate Gaussian distribution,

$$p(x) = \det(2\pi \mathbf{P}_x)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - \bar{x})^T \mathbf{P}_x^{-1}(x - \bar{x})\right\} \tag{18}$$

where $x \in \Re^{n \times 1}$ and $\mathbf{P}_x$ is a n by n covariance matrix, the above sigma points can be extended to the following form.

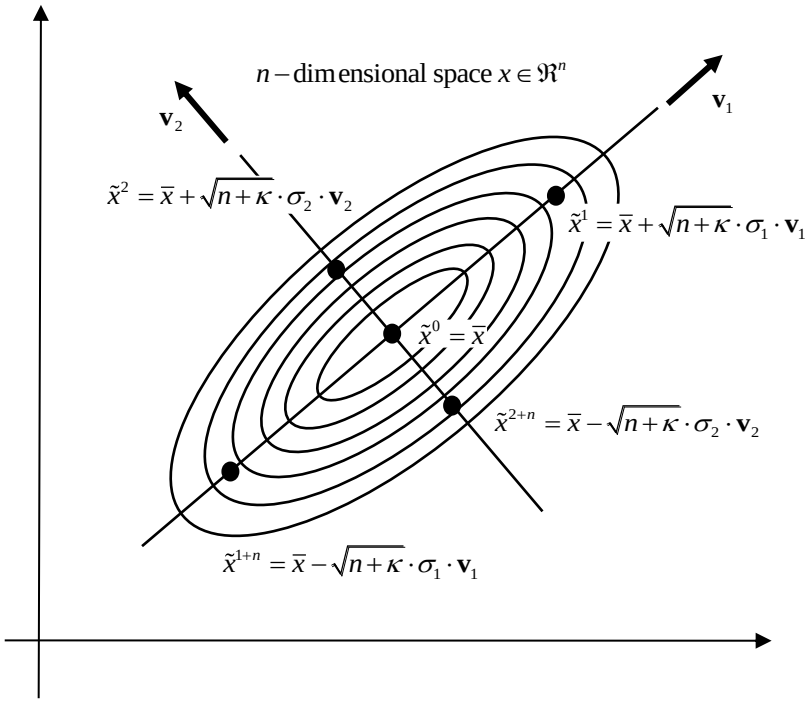


Figure 7.7 Sigma points for *n*-dimensional Gaussian distribution

Since the covariant matrix is a real, positive semi-definite matrix with non-negative eigenvalues, $\sigma_i^2 \geq 0, i = 1\cdots n$, it can be decomposed to:

$$\mathbf{P}_x = \mathbf{V}\mathbf{D}\mathbf{V}^T \tag{19}$$

where

$$\mathbf{V} = (\mathbf{v}_1 \cdots \mathbf{v}_n),\ \mathbf{D} = diag(\sigma_1^2 \cdots \sigma_n^2) \tag{20}$$

and $\mathbf{v}_i$ is the *i-th* unit eigenvector. Using these unit eigenvectors and eigenvalues, we can define (2*n*+1) sigma points in the *n* dimensional space.

$$\begin{aligned} &\tilde{x}^0 = \bar{x} \\ &\tilde{x}^i = \bar{x} + \sqrt{n+\kappa}\cdot\sigma_i\cdot\mathbf{v}_i \\ &\tilde{x}^{i+n} = \bar{x} - \sqrt{n+\kappa}\cdot\sigma_i\cdot\mathbf{v}_i \\ &\quad i = 1,\cdots,n \end{aligned} \tag{21}$$

with associated weights:

$$\begin{aligned} &W_0 = \frac{\kappa}{n+\kappa} \\ &W_i = W_{i+n} = \frac{1}{2(n+\kappa)} \\ &\quad i = 1,\cdots,n \end{aligned} \tag{22}$$

See Figure 7.7. As before let us propagate the (2*n*+1) sigma points through the

nonlinear function to obtain (2*n*+1) sample points in *y* that correspond to the individual sigma points in *x*.

$$\tilde{y}^i = f(\tilde{x}^i), \quad i = 0,1,\cdots,2n \tag{23}$$

For these propagated sample points $\{\tilde{y}_i \mid i = 0,1,\cdots 2n\}$, compute the weighted mean and covariance:

$$\bar{y}_{sample} = \sum_{i=0}^{2n} W_i \tilde{y}^i$$

$$P_{y,sample} = \sum_{i=0}^{2n} W_i (\tilde{y}^i - \bar{y}_{sample})(\tilde{y} - \bar{y}_{sample})^T \tag{24}$$

Again by setting $\kappa = 2$ we can show that these sample mean and sample covariance can approximate the true mean and covariance at least up to the second order.

$$\bar{y}_{sample} = E[y], \quad P_{y,sample} = E[(y - E[y])(y - E[y])^T]; \quad \kappa = 2 \tag{25}$$

This sampling method is called Unscented Transform.

7.5 Unscented Kalman Filter

Now that the unscented transform can approximate the mean and covariance up to the second order, let us apply it to Kalman filter. Consider a discrete-time, nonlinear state transition equation:

$$x_{t+1} = f(x_t, u_t, t) + w_t \tag{26}$$

and a nonlinear observation equation:

$$y_t = h(x_t, t) + v_t \tag{27}$$

where w_t and v_t are zero-mean, uncorrelated, Gaussian process noise and measurement noise, respectively.

Let us sample the sigma points from x_{t-1} as:

$$\begin{aligned} \tilde{x}_{t-1}^0 &= \hat{x}_{t-1} \\ \tilde{x}_{t-1}^i &= \hat{x}_{t-1} + \sqrt{n+\kappa} \cdot \sigma_i \cdot \mathbf{v}_i \\ \tilde{x}_{t-1}^{i+n} &= \hat{x}_{t-1} - \sqrt{n+\kappa} \cdot \sigma_i \cdot \mathbf{v}_i \\ & i = 1,\cdots,n \end{aligned} \tag{28}$$

where the unit eigenvectors and eigenvalues are those of the a posteriori error covariance:

$$P_{t-1} = E[(\hat{x}_{t-1} - x_{t-1})(\hat{x}_{t-1} - x_{t-1})^T] \tag{29}$$

Propagation:

Let $\tilde{x}^{i*}_{t|t-1}$ be the transformed sigma point of the *i-th* sigma point in the state space through the deterministic part of the nonlinear state transition equation.

$$\tilde{x}^{i*}_{t|t-1} = f(\tilde{x}^{i}_{t-1}, u_{t-1}, t-1), \quad i = 0, \cdots, 2n \tag{30}$$

The predicted mean of the a priori state estimation based on the (2n+1) samples is given by

$$\bar{x}_{t|t-1,sample} = \sum_{i=0}^{2n} W_i \tilde{x}^{i*}_{t|t-1} \tag{31}$$

and the predicted covariance is computed from (26) and (31) as

$$\begin{aligned} P_{t|t-1,sample} &= E[(\hat{x}_{t|t-1} - x_t)(\hat{x}_{t|t-1} - x_t)^T] \\ &= E[(\hat{x}_{t|t-1} - f(x_{t-1}, u_{t-1}, t-1) - w_{t-1})(\hat{x}_{t|t-1} - f(x_{t-1}, u_{t-1}, t-1) - w_{t-1})^T] \\ &= E[(\hat{x}_{t|t-1} - f(x_{t-1}, u_{t-1}, t-1))(\hat{x}_{t|t-1} - f(x_{t-1}, u_{t-1}, t-1))^T] + E[w_{t-1} {w_{t-1}}^T] \\ &= \sum_{i=0}^{2n} W_i (\tilde{x}^{i*}_{t|t-1} - \bar{x}_{t|t-1,sample})(\tilde{x}^{i*}_{t|t-1} - \bar{x}_{t|t-1,sample})^T + Q_{t-1} \end{aligned} \tag{32}$$

where the process noise w_{t-1} is uncorrelated with $x_{t-1}, \hat{x}_{t|t-1}$ and $f(x_{t-1}, u_{t-1}, t-1)$ is approximated to $\bar{x}_{t|t-1,sample}$. Note that the predicted mean and covariance are correct up to the second order. For brevity, the subscript "sample" will be dropped hereafter.

Update:

Similarly, a new set of sigma points $\tilde{x}^{i}_{t|t-1}, i = 0, \cdots, 2n$ are sampled for representing the distribution of $x_{t|t-1}$ by computing the eigenvalues and eigenvectors of $P_{t|t-1}$. These sigma points are mapped to output observations through the deterministic part of the observation equation:

$$\tilde{y}^{i}_{t} = h(\tilde{x}^{i}_{t|t-1}, t), \quad i = 0, \cdots, 2n \tag{33}$$

The predicted observation is computed as

$$\hat{y}_t = E[y_t \mid \hat{x}_{t|t-1}] \cong \sum_{i=0}^{2n} W_i \tilde{y}_t^i \tag{34}$$

The Kalman gain, too, can be computed from the samples. To this end, consider the following output estimation covariance, called the Innovation covariance,

$$\begin{aligned} P_y &\triangleq E\left[(y_t - \hat{y}_t)(y_t - \hat{y}_t)^T\right] \\ &= E\left[(H_t x_t + v_t - H_t \hat{x}_{t|t-1})(H_t x_t + v_t - H_t \hat{x}_{t|t-1})^T\right] \\ &= H_t E\left[(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^T\right] H_t^T + E\left[v_t v_t^T\right] \\ &= H_t P_{t|t-1} H_t^T + R_t \end{aligned} \tag{35}$$

Note that the inverse of this innovation covariance, P_y^{-1}, is involved in the Kalman gain, (5-38). The rest of the Kalman gain is computed as

$$\begin{aligned} P_{xy} &\triangleq E\left[(\hat{x}_{t|t-1} - x_t)(\hat{y}_t - y_t)^T\right] \\ &= E\left[(\hat{x}_{t|t-1} - x_t)(H_t \hat{x}_{t|t-1} - H_t x_t - v_t)^T\right] \\ &= E\left[(\hat{x}_{t|t-1} - x_t)(\hat{x}_{t|t-1} - x_t)^T\right] H_t^T - \underbrace{E\left[(\hat{x}_{t|t-1} - x_t) v_t^T\right]}_{0} \\ &= P_{t|t-1} H_t^T \end{aligned} \tag{36}$$

This is called the cross correlation matrix. Therefore, the Kalman gain is given by

$$K_t = P_{xy} P_y^{-1} \tag{37}$$

Both innovation covariance and cross correlation matrices can be evaluated by using the sigma points. The true output y_t is not available, but from (34) it can be written as

$$y_t = E[y_t] + v_t \cong \sum_{i=0}^{2n} W_i \tilde{y}_t^i + v_t \tag{38}$$

Using this in (35)

$$P_y = \sum_{i=0}^{2n} W_i (\tilde{y}_t^i - \hat{y}_t)(\tilde{y}_t^i - \hat{y}_t)^T + R_t \tag{39}$$

Similarly, the cross covariance is evaluated as

$$P_{xy} = \sum_{i=0}^{2n} W_i (\tilde{x}_{t-1}^i - \hat{x}_{t|t-1})(\tilde{y}_t^i - \hat{y}_t)^T \tag{40}$$

Covariance Update:

Recall $P_t = (I - K_t H_t) P_{t|t-1}$, this covariance update law can be rewritten without using H_t.

$$P_t = (I - K_t H_t) P_{t|t-1} = P_{t|t-1} - K_t H_t P_{t|t-1} = P_{t|t-1} - K_t P_y P_y^{-1} H_t P_{t|t-1} \tag{41}$$

Where $P_y P_y^{-1} = I$ is inserted in the last expression. Since $K_t = P_{t|t-1} H_t^T P_y^{-1}$,

$$P_t = P_{t|t-1} - K_t P_y K_t^T \tag{42}$$

From eqs. (32), (37), and (39), this can be computed using the sample points.

$$P_{t,sample} \cong P_{t|t-1,sample} - K_t P_y K_t^T \tag{43}$$

The recursive computation of Unscented Kalman Filter is summarized below. For brevity, subscript sample is omitted.

a) Given $\hat{x}_{t-1}$ and P_{t-1}, sample sigma points by computing eigenvalues and eigenvectors of P_{t-1}, (28),

b) Propagate the sigma points through the nonlinear model to obtain $\tilde{x}^{i*}_{t|t-1}$, (30),

c) From the (2*n*+1) propagated sigma points compute the mean and covariance, $\hat{x}_{t|t-1}$, $P_{t|t-1}$, (31) and (32),

d) Sample again (2*n*+1) sigma points $\tilde{x}^i_{t|t-1}$ for $P_{t|t-1}$,

e) Transform the propagated sigma points to output estimate $\tilde{y}^i_t$ based on the nonlinear measurement equation, (33), and compute the estimated output $\hat{y}_t$, (34),

f) Evaluate the innovation covariance and the cross covariance, P_y and P_{xy}, (39) and (40), by using (2n+1) samples of propagated output estimates,

g) Update the state estimation with the Kalman gain given by (37):

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t (y_t - \hat{y}_t)$$

h) Update the a posteriori covariance P_t using (43). Set $t = t+1$, and repeat the process.