4.7 Nonlinear State Estimation

In this section, we address the problem of state estimation for nonlinear stochastic systems. Let us consider the system

$$\begin{cases} x(t+1) = f\left(x(t), u(t), w(t)\right) \\ y(t) = h\left(x(t)\right) + v(t) \end{cases} \tag{4.70}$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ is a known deterministic input, $y(t) \in \mathbb{R}^p$ is the output, $w(t) \in \mathbb{R}^d$ is the process disturbance and $v(t) \in \mathbb{R}^p$ is the measurement noise. The functions $f : \quad \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^n$ and $h : \quad \mathbb{R}^n \to \mathbb{R}^p$ are assumed to be known, continuous and differentiable. For the process disturbance $w(t)$, measurement noise $v(t)$ and initial state $x(0)$ we enforce the same Assumption 4.1 as in the linear setting.

The problem we want to solve is the same as Problem 4.1: find an estimate of $x(t)$ based on input-output measurements up to time t. As we know, this is a Beyesian estimation problem, being both the unknown vector $x(t)$ and the data Y^t random variables. With respect to the linear case, the main difficulty lies in the fact that the functions $f(\cdot)$ and $h(\cdot)$ can be any $\mathcal{C}^1$ nonlinear mapping. Therefore, even if we restrict our attention to the LMSE estimate, the relevant covariance matrices at time t cannot be expressed only in terms of covariance matrices at previous times, because they will depend also on higher order moments (through the nonlinear model functions).

Since the exact solution of the LMSE problem is intractable, a wide variety of approximations have been proposed in the literature. It is important to keep in mind that all these techniques do not provide the actual LMSE state estimate, and hence they must be tested in simulations and real-world experiments in order to assess their performance in the specific application at hand. The most popular approach is the one based on the linearization of the model equations, which is known as *Extended Kalman Filter (EKF)*.

4.7.1 The Extended Kalman Filter

The EKF is a recursive procedure based on the same prediction-correction approach adopted in the linear KF. The notation adopted is the same as in Section 4.3, although both $\hat{x}(t|t)$ and $P(t|t)$ will end up to be only *approximations* of the true LMSE state estimate and covariance of the estimation error, respectively. We start by deriving the equations of the prediction step.

4.7.2 The EKF prediction step

Let $\hat{x}(t|t)$ and $P(t|t)$ be given. The objective is to compute the 1-step ahead prediction $\hat{x}(t+1|t)$ and the corresponding error covariance $P(t+1|t)$. To do this, we write the Taylor expansion of $f(x,u,w)$ around the nominal values $x = \hat{x}(t|t)$, $u = u(t)$ (recall that $u(t)$ is known at every time t) and $w = 0$ (being $\mathbf{E}[w(t)] = 0$). Then, one has

$$\begin{aligned} f(x,u,w) =& f(\hat{x}(t|t),u(t),0) + \left.\frac{\partial f}{\partial x}\right|_{\substack{x=\hat{x}(t|t)\\u=u(t)\\w=0}} (x-\hat{x}(t|t)) + \left.\frac{\partial f}{\partial u}\right|_{\substack{x=\hat{x}(t|t)\\u=u(t)\\w=0}} (u-u(t)) \\ &+ \left.\frac{\partial f}{\partial w}\right|_{\substack{x=\hat{x}(t|t)\\u=u(t)\\w=0}} (w-0) + O\left(\left|\left|\begin{smallmatrix} x-\hat{x}(t|t)\\ u-u(t)\\ w-0\end{smallmatrix}\right|\right|^2\right) \end{aligned}$$

where the last term contains all the terms of degree equal or higher than 2 in the involved variables. The main idea is that, *if these terms are small*, they can be neglected with respect to the linear terms. Hence, by doing so, we obtain the following approximated version of the first equation in (4.70)

$$x(t+1) \simeq f(\hat{x}(t|t),u(t),0) + \left.\frac{\partial f}{\partial x}\right|_{\substack{x=\hat{x}(t|t)\\u=u(t)\\w=0}} (x(t)-\hat{x}(t|t)) + \left.\frac{\partial f}{\partial w}\right|_{\substack{x=\hat{x}(t|t)\\u=u(t)\\w=0}} w(t) \tag{4.71}$$

In order to obtain $\hat{x}(t+1|t)$, we have to replace $x(t)$ and $w(t)$ in (4.71) with their best estimates based on the data available up to time t, which are clearly $\hat{x}(t|t)$ and 0, respectively (recall that being $w(t)$ white, its best

prediction is equal to the a priori expected value). Hence, one gets

$$\hat{x}(t+1|t) = f(\hat{x}(t|t), u(t), 0) \tag{4.72}$$

which turns out to be the first equation of the EKF prediction step. In order to update the covariance matrix, let us introduce the following notation for the Jacobian matrices of $f(\cdot)$ with respect to x and w

$$F(t) = \left.\frac{\partial f}{\partial x}\right|_{\substack{x=\hat{x}(t|t)\\ u=u(t)\\ w=0}} \in \mathbb{R}^{n\times n}$$

$$G(t) = \left.\frac{\partial f}{\partial w}\right|_{\substack{x=\hat{x}(t|t)\\ u=u(t)\\ w=0}} \in \mathbb{R}^{n\times d}$$

Then, by exploiting again (4.71) and (4.72), one gets

$$\begin{aligned}
P(t+1|t) &= E\left[(x(t+1) - \hat{x}(t+1|t))\,(x(t+1) - \hat{x}(t+1|t))^T\right] \\
&\simeq E\left[\{f(\hat{x}(t|t), u(t), 0) + F(t)(x(t) - \hat{x}(t|t)) + G(t)w(t)\right. \\
&\qquad \left. - f(\hat{x}(t|t), u(t), 0)\}\,\{\cdots\}^T\right] = \\
&= E\left[\{F(t)(x(t) - \hat{x}(t|t)) + G(t)w(t)\}\,\{\cdots\}^T\right] = \\
&= F(t)P(t|t)F(t)^T + G(t)QG(t)^T.
\end{aligned}$$

4.7.3 The EKF correction step

Let $\hat{x}(t+1|t), P(t+1|t)$ and $y(t+1)$ be given. Recall that

$$y(t+1) = h(x(t+1)) + v(t+1) \tag{4.73}$$

Let us write the Taylor expansion of $h(x)$ around the nominal value $x = \hat{x}(t+1|t)$ (the best available guess of $x(t+1)$ at time t). One has

$$\begin{aligned}
h(x) = h(\hat{x}(t+1|t)) &+ \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}(t+1|t)} (x - \hat{x}(t+1|t)) \\
&+ O\left(\left|\left|x - \hat{x}(t+1|t)\right|\right|^2\right)
\end{aligned} \tag{4.74}$$

where the last term contains all the terms of order equal or higher than 2 in the error $x - \hat{x}(t+1|t)$. As in the prediction step, we define the Jacobian matrix

$$H(t+1) = \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}(t+1|t)} \in \mathbb{R}^{p\times n}.$$

By using the expansion (4.74) into (4.73) and neglecting higher order terms, we get

$$y(t+1) \simeq h(\hat{x}(t+1|t)) + H(t+1)(x(t+1) - \hat{x}(t+1|t)) + v(t+1) \quad (4.75)$$

Now, denote the prediction error as

$$d(t+1) = x(t+1) - \hat{x}(t+1|t) \quad (4.76)$$

and set $m(t+1) = y(t+1) - h(\hat{x}(t+1|t))$. Then, (4.75) becomes

$$m(t+1) = H(t+1)d(t+1) + v(t+1). \quad (4.77)$$

By noticing that $\hat{d}(t+1|t) = \hat{x}(t+1|t) - \hat{x}(t+1|t) = 0$ and

$$\begin{aligned} &E\left[(d(t+1) - \hat{d}(t+1|t))(d(t+1) - \hat{d}(t+1|t))^T\right] \\ &= E\left[(x(t+1) - \hat{x}(t+1|t))(x(t+1) - \hat{x}(t+1|t))^T\right] = P(t+1|t) \end{aligned}$$

one can apply the correction step of the standard Kalman filter to the linear output equation (4.77) (observe that $m(t+1)$ is known). Therefore, one gets

$$\begin{aligned} \hat{d}(t+1|t+1) &= \hat{d}(t+1|t) + K(t+1)\left(m(t+1) - H(t+1)\hat{d}(t+1|t)\right) \\ &= K(t+1)(y(t+1) - h(\hat{x}(t+1|t))). \end{aligned}$$

From (4.76), one has $\hat{d}(t+1|t+1) = \hat{x}(t+1|t+1) - \hat{x}(t+1|t)$, which leads to the first EKF correction equation

$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t) + K(t+1)(y(t+1) - h(\hat{x}(t+1|t))) \quad (4.78)$$

where $K(t+1)$ is defined according to the standard KF applied to (4.77), i.e.

$$K(t+1) = P(t+1|t)H(t+1)^T\left[H(t+1)P(t+1|t)H(t+1)^T + R\right]^{-1} \quad (4.79)$$

and similarly one has that the update of the covariance matrix $P(t+1|t+1)$ is given by

$$P(t+1|t+1) = P(t+1|t)\left[I - H(t+1)^T K(t+1)^T\right]. \tag{4.80}$$

Hence, by assuming to initialize the EKF recursion in the same way as in the KF, the equations of the EKF algorithm can be summarized as follows.

Initialization: $\hat{x}(0|-1) = m_0,\ P(0|-1) = P_0$

For $t = 0, 1, 2, \cdots$

$$K(t) = P(t|t-1)H(t)^T[H(t)P(t|t-1)H(t)^T + R]^{-1} \tag{4.81}$$

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(t)(y(t) - h(\hat{x}(t|t-1))) \tag{4.82}$$

$$H(t) = \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}(t|t-1)} \tag{4.83}$$

$$P(t|t) = P(t|t-1)[I - H(t)^T K(t)^T] \tag{4.84}$$

$$\hat{x}(t+1|t) = f(\hat{x}(t|t), u(t), 0) \tag{4.85}$$

$$F(t) = \left.\frac{\partial f}{\partial x}\right|_{\substack{x=\hat{x}(t|t)\\u=u(t)\\w=0}}, \quad G(t) = \left.\frac{\partial f}{\partial w}\right|_{\substack{x=\hat{x}(t|t)\\u=u(t)\\w=0}} \tag{4.86}$$

$$P(t+1|t) = F(t)P(t|t)F(t)^T + G(t)QG(t)^T \tag{4.87}$$

end

4.7.4 Properties of the EKF

It is worth pointing out the main differences between the EKF and the KF derived for linear systems.

First, it is necessary to stress that the estimates $\hat{x}(t|t)$ and $\hat{x}(t+1|t)$ provided by the EKF are not the LMSE estimates of $x(t)$ and $x(t+1)$, respectively, based on Y^t. This is due to the approximations introduced in the linearization of the functions $f(x, u, w)$ and $h(x)$. Similarly, $P(t|t)$ and $P(t+1|t)$ are not the covariance matrices of the estimation errors, but only their approximations. How good such approximations are may depend on

several factors, including the initial conditions $\hat{x}(0|-1)$, $P(0|-1)$. Therefore, special care must be taken in the choice of such initial values, by exploiting the available a priori knowledge on the variables to be estimated.

Another main difference with the linear case is that the matrices $F(t)$, $G(t)$ and $H(t)$ depend on the current estimates $\hat{x}(t|t)$ and $\hat{x}(t+1|t)$ (that are the values at which the Jacobian matrice are computed), which in turn depend on the data Y^t. As a consequence, the matrices $K(t), P(t|t), P(t+1|t)$ cannot be precomputed as in the linear case. Moreover, they depend on the specific data realization processed, which means that also the uncertainty associated to the estimates actually depend on the data set. In other words, we cannot assess the quality of the estimates before computing them.

In general, there are no guarantees that the estimates provided by the EKF are satisfactory. Loosely speaking, neglecting the higher order terms is reasonable only if the estimation errors are "small", which clearly leads to a circular reasoning. The possible undesired behaviors that can be observed can be divided in two types: *divergence* occurs when the error grows arbitrarily, i.e.

$$\lim_{t\to\infty} ||x(t) - \hat{x}(t|t)|| = +\infty.$$

On the other hand, even if the error remains bounded, one may face *inconsistency* of the estimates. This happens if

$$E\left[(x(t) - \hat{x}(t|t))(x(t) - \hat{x}(t|t))^T\right] \gg P(t|t)$$

that is, if the actual uncertainty is much larger than the one evaluated by the EKF. In such a case, one may make overoptimistic statements about the quality of the estimates delivered by the filter. Said another way, the true state values may be significantly faraway from the confidence intervals derived from the filter estimates.

State estimation for nonlinear dynamic systems is still an active research area. Whenever the solution provided by the EKF is not satisfactory, one may resort to a variety of alternatives that have been proposed in the literature. Some of them are briefly presented in Sections 4.9 and 4.10.

4.8 The Continuous-Discrete Kalman Filter

In many real-world applications, the system dynamics are described by a set of continuous-time differential equations, while the output measurements are available only at discrete time instants. For example, this is the case of sampled-data systems, in which the physical behavior is usually described by a state-space continuous-time model, while the sensor measurements are functions of the state vector, sampled at a given clock rate or made available at asynchronous time instants. For a linear time-invariant system, this setting is captured by the equations

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + Gw(t) \\ y(t_k) = Cx(t_k) + v(t_k) \qquad k = 0, 1, 2, \dots \end{cases}$$

where $t_0 < t_1 < t_2 < \dots$ are the discrete time instants at which measurements are collected. It is worth observing that in this case $w(t)$ is a continuous-time stochastic process which is assumed to be stationary and white. This means that its covariance function is given by $\mathbf{E}\left[w(t+\tau)w^T(t)\right] = Q\delta(\tau)$, where Q is the spectral density matrix. We also set $\mathbf{E}\left[w(t)\right] = 0$ and make the usual assumptions on the output noise, i.e., $v(t_k) \sim WP\left(0, R\right)$ and it is independent from $w(t)$.

The LMSE state estimates for system (4.8) is given by the following recursive equations:

Initialization: $\hat{x}(t_0|t_{-1}) = m_0,\ P(t_0|t_{-1}) = P_0$

For $k = 0, 1, 2, \cdots$

$$K(t_k) = P(t_k|t_{k-1})C^T[CP(t_k|t_{k-1})C^T + R]^{-1} \tag{4.88}$$

$$\hat{x}(t_k|t_k) = \hat{x}(t_k|t_{k-1}) + K(t_k)(y(t_k) - C\hat{x}(t_k|t_{k-1})) \tag{4.89}$$

$$P(t_k|t_k) = P(t_k|t_{k-1})[I - C^T K(t_k)^T] \tag{4.90}$$

$$\frac{d}{dt}\hat{x}(t|t_k) = A\hat{x}(t|t_k) + Bu(t)\ , \qquad \text{for } t \in [t_k, t_{k+1}) \tag{4.91}$$

$$\frac{d}{dt}P(t|t_k) = AP(t|t_k) + P(t|t_k)A^T + GQG^T\ , \quad \text{for } t \in [t_k, t_{k+1}) \tag{4.92}$$

end

Algorithm (4.88)-(4.92) is known as the *Continuous-Discrete Kalman Filter (CDKF)*. While equations (4.88)-(4.90) are the same as in the standard discrete-time Kalman filter, the equations of the prediction step (4.91)-(4.92) are a set of differential equations which allow one to compute the prediction of the state and the associated covariance matrix of the prediction error along the time interval $t \in [t_k, t_{k+1})$, until the next measurements $y(t_{k+1})$ becomes available.

It is worth observing that (4.91) are n first-order linear differential equations in the state components, while (4.92) correspond to $\frac{n(n+1)}{2}$ first-order linear differential equations in the independent entries of $P(t|t_k)$ (recall that P is symmetric). Hence the solutions of these two systems of differential equations can be computed analytically. It must be also stressed that the resulting state estimate trajectories are piecewise continuous, as $\hat{x}(t|t_k)$ is continuous in the interval $t \in (t_k, t_{k+1})$, while jumps occur at the discrete time instants t_k due to the correction step (the same occurs also for the entries of P).

Clearly, a possible alternative to the CDKF described above is to first discretize the continuous-time dynamics of system (4.8) and then apply the discrete-time Kalman filter to the discretized system. This usually works well for linear systems, which can be discretized *exactly*, i.e., without introducing errors in the dynamics due to the discretization. On the other hand, the state estimation problem becomes much more challenging when dealing with nonlinear systems. In particular, if the time interval between two time samples t_k and t_{k+1} is long (compared to the time constants of the system dynamics), a poor discretization of the system dynamics may lead to significant errors, which in turn can generate inconsistency of the estimates or even divergence of the filter.

A possible solution of the state estimation problems in the case of sampled-data nonlinear systems is provided by the *Continuous-Discrete Extended Kalman Filter (CDEKF)* described next.

Consider the system

$$\begin{cases} \dot{x}(t) = f\left(x(t), u(t), w(t)\right) \\ y(t_k) = h\left(x(t_k)\right) + v(t_k) \end{cases} \tag{4.93}$$

in which $w(t)$ and $v(t_k)$ satisfy the same assumptions as in the linear case treated above. Then, the equations of the CDEKF algorithm are given by

Initialization: $\hat{x}(t_0|t_{-1}) = m_0,\ P(t_0|t_{-1}) = P_0$

For $k = 0, 1, 2, \cdots$

$$K(t_k) = P(t_k|t_{k-1})H^T(t_k)[H(t_k)P(t_k|t_{k-1})H^T(t_k) + R]^{-1} \tag{4.94}$$

$$\hat{x}(t_k|t_k) = \hat{x}(t_k|t_{k-1}) + K(t_k)(y(t_k) - h(\hat{x}(t_k|t_{k-1})) \tag{4.95}$$

$$P(t_k|t_k) = P(t_k|t_{k-1})[I - H^T(t_k)K(t_k)^T] \tag{4.96}$$

$$\frac{d}{dt}\hat{x}(t|t_k) = f(\hat{x}(t|t_k), u(t), 0)\ , \qquad \text{for } t \in [t_k, t_{k+1}) \tag{4.97}$$

$$\frac{d}{dt}P(t|t_k) = F(t_k)P(t|t_k) + P(t|t_k)F^T(t_k) + G(t_k)QG^T(t_k)\ , \qquad \text{for } t \in [t_k, t_{k+1}) \tag{4.98}$$

end

where

$$F(t_k) = \left.\frac{\partial f}{\partial x}\right|_{\substack{x=\hat{x}(t_k|t_k) \\ u=u(t_k) \\ w=0}} \in \mathbb{R}^{n\times n} \tag{4.99}$$

$$G(t_k) = \left.\frac{\partial f}{\partial w}\right|_{\substack{x=\hat{x}(t_k|t_k) \\ u=u(t_k) \\ w=0}} \in \mathbb{R}^{n\times d} \tag{4.100}$$

$$H(t_k) = \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}(t_k|t_k-1)} \in \mathbb{R}^{p\times n} \tag{4.101}$$

As in the linear case, the correction step composed by equations (4.94)-(4.96) are analogous to those of the discrete-time EKF. The prediction step (4.97)-(4.98) consists of n nonlinear differential equations in the state predictions $\hat{x}(t|t_k)$ and $\frac{n(n+1)}{2}$ linear differential equations in the entries of $P(t|t_k)$. Such

equations must be integrated over the interval $t \in [t_k, t_{k+1})$. In particular, the crucial task is the integration of the nonlinear equations (4.97). In fact, the main reason why the CDEKF is often successfully employed in applications involving sampled-data systems is that it allows one to precisely evolve the state estimates between two consecutive measurements, through precise numerical integration of the prediction step equations.

Clearly, the same observations made for the EKF in the discrete-time setting, apply also to the CDEKF. In particular, there is no guarantee that the expected value of the state estimation error converge to zero, nor that $P(t|t_k)$ be equal to the covariance matrix of the estimation error. In this respect, the role of the initial conditions $\hat{x}(t_0|t_{-1})$ and $P(t_0|t_{-1})$ may be crucial to promote a satisfactory behavior of the filter.

Remark 4.2. In principle, the Jacobian matrices (4.99) and (4.100) could be computed by linearizing the system around the current prediction $\hat{x}(t|t_k)$ and the current input $u(t)$, instead of their corresponding values at the beginning of the integration interval $[t_k, t_{k+1})$. Notice however that this would make the matrices F and G time-varying and, more importantly, dependent on the variables $\hat{x}(t|t_k)$ themselves! As a consequence, the equations (4.97)-(4.98) would be coupled, that is they would become a unique system of nonlinear differential equations in $n + \frac{n(n+1)}{2}$ variables. The resulting increase of the computational burden may not be worth the advantage provided by a more precise approximation of the covariance matrix of the estimation errors.

4.9 The Unscented Kalman Filter

The Unscented Kalman Filter (UKF) is an algorithm for state estimation of nonlinear systems, based on the *unscented transform*. The key idea is to use a set of points in the state space, in order do match some relevant statistics of the a posteriori pdf of the state $f_x(x(t)|Y^t)$. Such points are called *sigma points* and are denoted hereafter by $X^{(i)} \in \mathbb{R}^n$, for $i = 1, \ldots, p$. To each sigma point we associate a scalar weight $W^{(i)} \in \mathbb{R}$, $i = 1, \ldots, p$, such

that $\sum_{i=1}^{p} W^{(i)} = 1$. The sigma points are chosen in such a way that their (weighted) sample statistics are equal to the selected statistics of the target a posteriori pdf.

Example 4.1. Let $x \in \mathbb{R}^n$ be a Gaussian random variable, $x \sim N(m, P)$. Let us choose $p = 2n$ sigma points as

$$\begin{aligned} X^{(i)} &= m + (\sqrt{nP})_i \quad i = 1, \cdots, n \\ X^{(i)} &= m - (\sqrt{nP})_i \quad i = n+1, \cdots, 2n \\ W^{(i)} &= \frac{1}{2n} \quad i = 1, \cdots, 2n \end{aligned}$$

where $(M)_i$ denotes the i-th column of matrix M, while for M and U square matrices, $\sqrt{M} = U$ means that $M = UU^T$. It is easy to show that the weighted mean and covariance matrix of the sigma points match those of the original Gaussian pdf, i.e.

$$\begin{aligned} \sum_{i=1}^{2n} W^{(i)} X^{(i)} &= m, \\ \sum_{i=1}^{2n} W^{(i)} (X^{(i)} - m)(X^{(i)} - m)^T &= P. \end{aligned}$$

We aim to use the sigma points to propagate the relevant statistics of a random variable that undergoes a nonlinear transformation. Consider the non linear function $h : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$z = h(x)$$

Let $X^{(i)}, W^{(i)}$ be a set of sigma points for x. If we apply the nonlinear function $h(\cdot)$ to the sigma points, we obtain the new set of sigma points

$$Z^{(i)} = h(X^{(i)}) \quad i = 1, \cdots, p.$$

Then, we approximate the statistic of z with the new sigma points. For example, the mean

$$\mathbf{E}[z] \simeq \sum_{i=1}^{p} W^{(i)} Z^{(i)} = m_z$$

and the covariance matrix

$$\mathrm{Cov}(z) \simeq \sum_{i=1}^{p} W^{(i)}(Z^{(i)} - m_z)(Z^{(i)} - m_z)^T$$

The main idea behind the UKF algorithm is to use the sigma points to propagate the statistics of the posterior pdf $f_x(x(t)|Y^t)$, using a prediction-correction structure similar to that of the EKF.

4.9.1 The UKF Algorithm

Consider system (4.70) and let Assumption 4.1 hold. Let $\hat{x}(t|t)$ and $P(t|t)$ have the same meaning as in the EKF and assume they are given at time t. The prediction and correction step of the UKF are summarized next.

Prediction

- Generate the sigma points $\{X^{(i)}, W^{(i)}\}$ so that they have mean $\hat{x}(t|t)$ and covariance $P(t|t)$.
- Generate the sigma points $\{\Xi^{(i)}, W^{(i)}\}$ so that they match some relevant statistics of the pdf of $w(t)$ (e.g., the mean 0 and the covariance Q).
- Compute the predicted sigma points

$$\hat{X}^{(i)} = f(X^{(i)}, u(t), \Xi^{(i)}) \quad i = 1, \ldots, p$$

- Compute the (approximated) mean and covariance of the a posteriori pdf $f_x(x(t+1)|Y^t)$ as

$$\hat{x}(t+1|t) = \sum_{i=1}^{p} W^{(i)} \hat{X}^{(i)}$$

$$P(t+1|t) = \sum_{i=1}^{p} W^{(i)}(\hat{X}^{(i)} - \hat{x}(t+1|t))(\hat{X}^{(i)} - \hat{x}(t+1|t))^T$$

Correction

- Generate the sigma points $\{V^{(i)}, W^{(i)}\}$ so that they match some relevant statistics of the pdf of $v(t)$ (e.g., the mean 0 and the covariance R).

- Compute the output sigma points

$$\hat{Y}^{(i)} = h(\hat{X}^{(i)}) + V^{(i)} \quad i = 1, \cdots, p$$

- Set

$$\hat{y}(t+1) = \sum_{i=1}^{p} W^{(i)} \hat{Y}^{(i)}$$

$$S(t+1) = \sum_{i=1}^{p} W^{(i)} (\hat{Y}^{(i)} - \hat{y}(t+1))(\hat{Y}^{(i)} - \hat{y}(t+1))^T$$

$$P_{xy}(t+1) = \sum_{i=1}^{p} W^{(i)} (\hat{x}^{(i)} - \hat{x}(t+1|t))(\hat{Y}^{(i)} - \hat{y}(t+1))^T$$

- Compute the (approximated) mean and covariance of the a posteriori pdf $f_x(x(t+1)|Y^{t+1})$ as

$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t) + P_{xy}(t+1) S(t+1)^{-1} \left(y(t+1) - \hat{y}(t+1)\right)$$

$$P(t+1|t+1) = P(t+1|t) - P_{xy}(t+1) S(t+1)^{-1} P_{xy}(t+1)^T.$$

It is apparent that the last two equations derive from the LMSE estimation expressions, in which the relevant covariances are obtained from the corresponding sample statistics of the sigma points.

It is worth stressing that many different versions of the UKF have been proposed in the literature. The one presented above contains only the basic features, but the tool is flexible enough to allow several interesting extensions. For example, one may take into account cross-correlations between state variables, process disturbances and measurement noise, by generating a set of sigma points representative of the joint statistics of an extended vector including $x(t)$, $w(t)$ and $v(t)$. Moreover, one may exploit a priori knowledge on the functional form of the involved distributions and generate

and extended set of sigma points matching higher order statistics, beyond the mean and the covariance. For further details, see the references (Julier and Uhlmann, 2004; Wan *et al.*, 2001).

4.10 The Particle Filter

Ideally, the aim of the state estimation problem is to find an estimate of the a posteriori pdf of the state, $f_x(x(t)|Y^t)$. Then, one can compute the MSE estimate of $x(t)$ as

$$\mathbf{E}\left[x(t)|Y^t\right] = \int x f(x|Y^t) dx.$$

In the linear Gaussian case (i.e., if the system is linear and the stochastic processes $w(t)$ and $v(t)$ have a Gaussian pdf), it turns out that

$$f(x(t)|Y^t) = N(\hat{x}(t|t), P(t|t))$$

where $\hat{x}(t|t)$ and $P(t|t)$ can be computed through the KF iterations. If the Gaussian assumption is not satisfied, the KF still provides the LMSE estimate of the state. However, in the nonlinear case, the a posteriori pdf of $x(t)$ can be significantly different from a Gaussian pdf. In such cases, the pdf $N(\hat{x}(t|t), P(t|t))$, with $\hat{x}(t|t)$ and $P(t|t)$ provided by the EKF, can be a very coarse approximation of the true pdf $f_x(x(t)|Y^t)$.

In principle, one can update the a posteriori pdf of the state through a prediction-correction iterative method, as explained next. Consider again system (4.70) and assume that $w(t)$ and $v(t)$ are independent and distributed according to $f_w(w(t))$ and $f_v(v(t))$, respectively.

Prediction

Assume that $f_x(x(t)|Y^t)$ is known. From the theory of joint and conditional pdfs, one has

$$\begin{aligned} f_x(x(t+1)) &= \int f_x(x(t+1), x(t)) dx(t) \\ &= \int f_x(x(t+1)|x(t)) f_x(x(t)) dx(t) \end{aligned}$$

The above equation still holds if we condition all the involved pdfs also to the data Y^t, thus giving

$$\begin{aligned} f_x(x(t+1)|Y^t) &= \int f_x(x(t+1)|x(t),Y^t) f_x(x(t)|Y^t) dx(t) \\ &= \int f_x(x(t+1)|x(t)) f_x(x(t)|Y^t) dx(t) \end{aligned} \tag{4.102}$$

where the last equality is due to the fact that $x(t)$ is a Markov process. Therefore, in order to compute the predicted pdf $f_x(x(t+1)|Y^t)$, we need to know $f_x(x(t+1)|x(t))$, which in turn can be derived from the first equation in model (4.70) and the knowledge of $f_w(w(t))$.

Correction

From Bayes's formula, one has

$$\begin{aligned} f_x(x(t+1)|Y^{t+1}) &= f_x(x(t+1)|y(t+1),Y^t) \\ &= \frac{f_{xy}(x(t+1),y(t+1)|Y^t)}{f_y(y(t+1)|Y^t)} \\ &= \frac{f_y(y(t+1)|x(t+1)) f_x(x(t+1)|Y^t)}{f_y(y(t+1)|Y^t)} \end{aligned} \tag{4.103}$$

where in the last equality we have exploited the fact that $f_y(y(t+1)|x(t+1),Y^t) = f_y(y(t+1)|x(t+1))$, thanks once again to the fact that $x(t)$ is a Markov process. Now, being

$$\begin{aligned} f_y(y(t+1)|Y^t) &= \int f_y(y(t+1)|x(t+1),Y^t) f_x(x(t+1)|Y^t) dx(t+1) \\ &= \int f_y(y(t+1)|x(t+1)) f_x(x(t+1)|Y^t) dx(t+1) \end{aligned}$$

by substituting in (4.103) one gets

$$\begin{aligned} &f_x(x(t+1)|Y^{t+1}) = \\ &= \frac{f_y(y(t+1)|x(t+1))}{\int f_y(y(t+1)|x(t+1)) f_x(x(t+1)|Y^t) dx(t+1)} f_x(x(t+1)|Y^t). \end{aligned} \tag{4.104}$$

Therefore, in order to compute the corrected pdf $f_x(x(t+1)|Y^{t+1})$, we need to know $f_y(y(t+1)|x(t+1))$, which can be derived from the second equation in model (4.70) and the knowledge of $f_v(v(t))$. In particular, one has

$$f_y(y(t+1)|x(t+1)) = f_v(y(t+1) - h(x(t+1)). \tag{4.105}$$

The equations (4.102) and (4.104) can be used in principle to compute the exact a posteriori pdf of the state vector. However, such computations can be intractable even by numerical approximation techniques, especially when the dimension of the state vector is large. One way to practically approximate the a posterior pdf is through the so-called Monte Carlo sequential methods. The key idea is to use a relatively large set of points in the state space ("particles") whose sample distribution is as close as possible to the true a posteriori pdf $f_x(x(t)|Y^t)$. The resulting estimator, known as *particle filter*, is described next.

4.10.1 The Particle Filter Algorithm

Consider system (4.70) and let $f_w(w(t))$ and $f_v(v(t))$ be known. Assume that a set of particles x_t^i, for $i = 1, \cdots, N$, is available and that they are (approximately) distributed according to $f_x(x(t)|Y^t)$.

Prediction

- Generate N particles w_t^i, $i = 1, \cdots, N$, distributed according to $f_w(w(t))$.
- Compute the predicted particles

$$\hat{x}_{t+1}^i = f(x_t^i, u(t), w_t^i) \qquad i = 1, \cdots, N$$

  The particles $\hat{x}_{t+1}^i$ provide an approximation of $f_x(x(t+1)|Y^t)$.

Correction

- Generate the weights

$$q_i = \frac{f_y(y(t+1)|\hat{x}_{t+1}^i)}{\sum_{j=1}^{N} f_y(y(t+1)|\hat{x}_{t+1}^j)} \qquad i = 1, \cdots, N$$

  One clearly has $q_i \geq 0$, $\forall i$, and $\sum_{i=1}^{N} q_i = 1$. We use q_i as the probability mass function of $\hat{x}_{t+1}^i$ in the a posteriori pdf $f_x(x(t+1)|Y^{t+1})$. Notice that, according to (4.105), $f_y(y(t+1)|\hat{x}_{t+1}^i) = f_v(y(t+1) - h(\hat{x}_{t+1}^i))$.

- *Re-sampling.* Sample N-times from the set of particles $\hat{x}^i_{t+1}$ in such a way that the probability of extracting $\hat{x}^i_{t+1}$ is equal to q_i. This amounts to generate a new set of particles x^j_{t+1}, $j = 1, \ldots, N$, such that

$$P\{x^j_{t+1} = \hat{x}^i_{t+1}\} = q_i \ , \qquad j = 1, \cdots, N.$$

  The new set of particles x^j_{t+1}, $j = 1, \cdots, N$, is approximately distributed according to $f_x(x(t+1)|Y^{t+1})$.

As for the UKF, the above one is a basic version of the particle filter. A number of variations have been proposed in the literature which allow the user to exploit the potential of this approach to carefully approximate probability density functions even in state spaces of very high dimensions. We also hint to the fact that there are theoretical results showing that, under suitable technical assumptions, by increasing the number of particles N one can approximate the true pdf $f_x(x(t)|Y^t)$ with arbitrary precision. For a detailed treatment of the subject, see the references (Gordon *et al.*, 1993; Doucet *et al.*, 2001).

Bibliography

Doucet, Arnaud, Nando De Freitas and Neil Gordon (2001). An introduction to sequential Monte Carlo methods. In: *Sequential Monte Carlo methods in practice*. pp. 3–14. Springer.

Gordon, Neil J., David J. Salmond and Adrian F.M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEE proceedings F (radar and signal processing)*. Vol. 140. IET. pp. 107–113.

Julier, Simon J. and Jeffrey K. Uhlmann (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE* **92**(3), 401–422.

Wan, Eric A., Rudolph Van Der Merwe and Simon Haykin (2001). The unscented Kalman filter. *Kalman filtering and neural networks* **5**(2007), 221–280.