Chapter 4

Introduction to Bayesian State Estimation

4.1 Preliminaries

Lower-case boldface letters $\boldsymbol{x}$ denote a random variable, whose probability distribution is $f(x)$. An underlined quantity, such as $\underline{x}$, denotes a vector. The notation $\hat{x}$ is used to describe either the (1) mean, (2) the estimate or (3) the observed value of a random variable $\boldsymbol{x}$. Matrices are written in upper-case bold letters.

4.2 Motivation

Stochastic methods for state estimation can lead to better results when compared to previous methods. In previous methods, the position of a tag was solved for a current set of measurements without any prior information. Realistically, measurements can be noisy due to the limitations of a given sensor. In Section 3.2, several sources of error are outlined that illustrate why UWB measurements can differ considerably from the true value despite its technological advantages. Moreover, it is also possible that several measurements do not arrive at the computing node due to packet loss, faulty hardware or other communication issues. This could lead to an under-determined system of equations that cannot be solved by geometric methods. Stochastic methods aim to ameliorate these problems by characterizing this measurement uncertainty and making use of a system model.

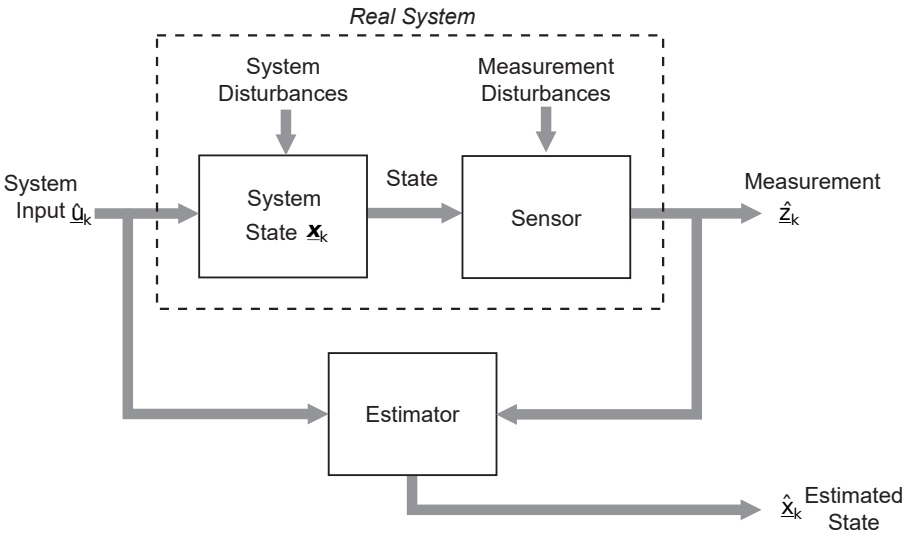

Figure 4.1: The estimation of a real state $\underline{\boldsymbol{x}}$ by a state estimator. The estimator combines noisy system inputs $\hat{\underline{u}}$ and measurements $\hat{\underline{z}}$ in order to construct the estimated state $\hat{\underline{x}}$. Image source: Hanebeck et al. [35].

If the behaviour of a state, $\underline{\boldsymbol{x}}$, can be modelled, the state transition model denoted by the function $\underline{a}_k = (.,.)$[1] can map the current state $\underline{\boldsymbol{x}}_k$ to the next state $\underline{\boldsymbol{x}}_{k+1}$ as

$$\underline{\boldsymbol{x}}_{k+1} = \underline{a}_k(\underline{\boldsymbol{x}}_k, \hat{\underline{u}}_k, \underline{\boldsymbol{w}}_k), \tag{4.1}$$

where $\hat{\underline{u}}_k$ is the vector of inputs and $\underline{\boldsymbol{w}}_k$ accounts for the uncertainty of the system model at time-step k [35]. As explained in [35], it is important to note that no model can perfectly describe the behaviour of a system. For example, the wheels of a vehicle are prone to slipping, its movement can be affected by wind and so forth. These factors can be too complex to model deterministically, which is the motivation for describing this uncertainty with a stochastic component $\underline{\boldsymbol{w}}_k$.

A measurement model can be described by the non-linear function

$$\underline{\boldsymbol{z}}_k = \underline{h}_k\left(\underline{\boldsymbol{x}}_k, \underline{\boldsymbol{v}}_k\right), \tag{4.2}$$

which maps the current system state $\underline{\boldsymbol{x}}_k$ to the measured quantity $\hat{\underline{z}}_k$. The measurement $\hat{\underline{z}}$ is the realisation, or, in other words, the obtained value by a sensor, of the random measurement vector $\underline{\boldsymbol{z}}$. Similar to the system model, an additional stochastic component, $\underline{\boldsymbol{v}}_k$, can describe the uncertainty of the measurement model. These two models are shown in Figure 4.2.

[1] In the general case, the system model $\underline{a}_k$ is described as non-linear and time-variant. Nevertheless, some systems can be described using a linear time-invariant system model. The same also applies to the non-linear measurement model.

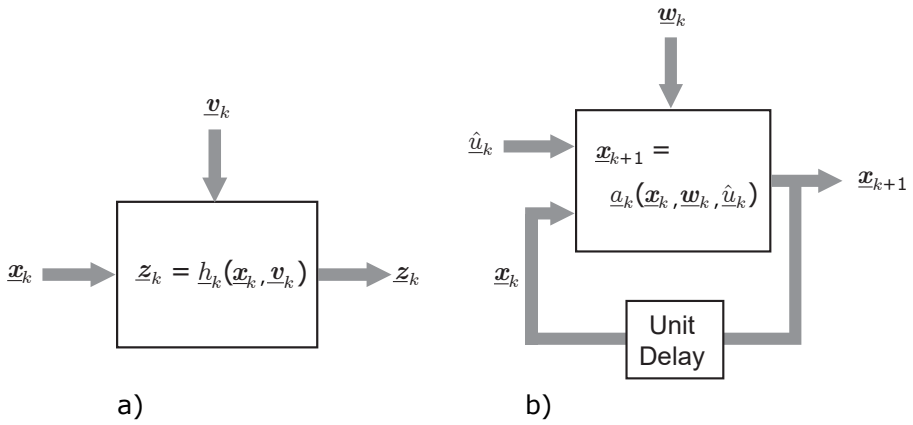

Figure 4.2: A stochastic measurement (a) and system model (b). Image source: Hanebeck et al. [35].

From the system model, it is possible to *predict* the next state, $\underline{\boldsymbol{x}}_{k+1}^{\mathrm{p}}$, from the current estimate. Doing this iteratively over multiple time-steps would increase the uncertainty of the system due to inaccuracies in the system model. Therefore, measurements from the current time-step are used to *filter* this predicted state $\underline{\boldsymbol{x}}_{k+1}^{\mathrm{p}}$. This process, known as *recursive filtering* [65], characterizes the Bayesian filters that are outlined in this chapter. The aforementioned issues, namely insufficient measurements and measurement noise, are mitigated by utilizing the prediction step to construct a state in the current time-step, even if no measurements are received and by using stochastic terms to describe the measurement and state uncertainty in order to decide how the measurements filter the predicted state.

4.3 Bayesian State Estimation

A Bayesian estimator recursively estimates the probability density of a random vector $f^{e}(\underline{x})$. In order to achieve this, the system input $\hat{\underline{u}}$ and the measurement $\hat{\underline{z}}$, which is the realisation of $\underline{\boldsymbol{z}}$ (see Section 4.2), are used. The probability density depicting the prior knowledge at time step $k = 0$ is described as $f^{\mathrm{e}}(\underline{x}_0)$ and is assumed to be known[2].

Prediction

Beginning from the initial time-step, we regard a time-step k. The prior density known by the Bayesian estimator can be described as

$$f^{\mathrm{e}}(\underline{x}_k) = f(\underline{x}_k | \hat{\underline{z}}_{0:k}, \hat{\underline{u}}_{0:k-1}), \tag{4.3}$$

[2]In reality, the prior probability density $f^{\mathrm{e}}(\underline{x}_0)$ is conservatively defined to account for the uncertainty.

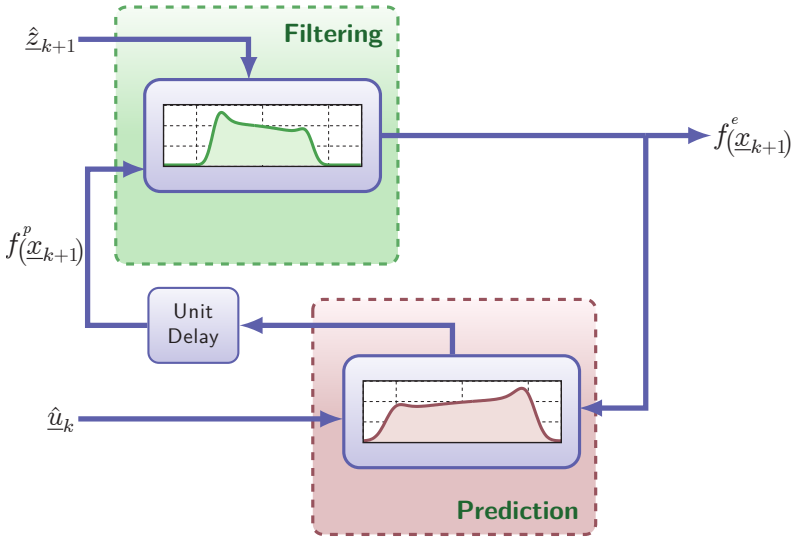

Figure 4.3: Bayesian estimation consisting of prediction and filtering steps. Image source: Noack [62].

which describes the probability density $f^{\mathrm{e}}(\underline{x}_k)$ conditioned the measurements $\hat{\underline{z}}_{0:k}$ and the system inputs $\hat{\underline{u}}_{0:k-1}$. Bayesian filters assume that the dynamic system is Markovian, which means that the current state, described by $\underline{x}_k$ and its density $f^{\mathrm{e}}(\underline{x}_k)$, provides as much information about the system as all states prior to and including the current state, $\underline{x}_{0:k}, \{f^{\mathrm{e}}(\underline{x}_0), \ldots, f^{\mathrm{e}}(\underline{x}_k)\}$. This allows the states to be treated as conditionally independent. Therefore, only the current state needs to be used for predicting the next state.

The state transition model $\underline{a}_k(\,\cdot\,)$, shown in (4.1), can deterministically predict $\underline{x}_{k+1}$ using the current input, $\hat{\underline{u}}_k$ and the current state vector $\underline{x}_k^{\mathrm{k}}$ for a fixed value, $\underline{w}_k$, of $\underline{\boldsymbol{w}}_k$. As shown in [62], the predicted probability density $f^{\mathrm{p}}(\underline{x}_{k+1})$ is provided by the Chapman-Kolmogorov integral

$$f^{\mathrm{p}}(\underline{x}_{k+1}) = \int_{\mathbb{R}_x^n} f\left(\underline{x}_{k+1}|\underline{x}_k, \hat{\underline{u}}_k\right) \cdot f^{\mathrm{e}}\left(\underline{x}_k\right) \mathrm{d}\underline{x}_k \tag{4.4}$$

In general, a joint density can be marginalized [11] as follows

$$f\left(x|y\right) = \int_Z f\left(x, z|y\right) \mathrm{d}z. \tag{4.5}$$

Applying this rule allows us to marginalize $\underline{x}_k$ from the Chapman-Kolmogorov integral (4.4) to obtain the predicted density

$$f^{\mathrm{p}}\left(\underline{x}_{k+1}\right) = f\left(\underline{x}_{k+1}|\hat{\underline{u}}_{0:k}, \hat{\underline{z}}_{0:k}\right). \tag{4.6}$$

Filtering

Once the measurement for the current time-step $\hat{\underline{z}}_{k+1}$ arrives, Bayes' rule can be applied. Bayes' rule states

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{4.7}$$

By applying this rule, the posterior $f^{\mathrm{e}}\left(\underline{x}_{k+1}\right)$, can be determined by

$$f^{\mathrm{e}}\left(\underline{x}_{k+1}\right) = \frac{f\left(\hat{\underline{z}}_{k+1}|\underline{x}_{k+1}\right) \cdot f^{\mathrm{p}}\left(\underline{x}_{k+1}\right)}{\int_{\mathbb{R}^{n_x}} f\left(\hat{\underline{z}}_{k+1}|\underline{x}_{k+1}\right) \cdot f^{\mathrm{p}}\left(\underline{x}_{k+1}\right) \mathrm{d}\underline{x}_k}, \tag{4.8}$$

where the denominator is a normalization constant. The density $f\left(\hat{\underline{z}}_{k+1}|\underline{x}_{k+1}\right)$ can be determined by the measurement model.

4.4 Kalman Filter

Though the general Bayesian recursive state estimator provides an optimal solution by minimizing the mean square error (MSE), it is unfortunately impractical to implement because the probability densities cannot be computed efficiently in closed-form due to the infinitely large parametrizations [62, 83]. Proposed by Rudolf. E. Kalman in 1960 [42], the Kalman Filter provides an optimal solution to this problem if certain conditions are met [62, 75], namely:

1. The probability density of $\underline{\boldsymbol{x}}$ is Gaussian. Therefore, $\underline{\boldsymbol{x}} \sim \mathcal{N}(\underline{\hat{x}}, \mathbf{C})$, where $\underline{\hat{x}}$ is the mean and $\mathbf{C}$ denotes the covariance matrix, which describes the variances of the components of the vector $\underline{\boldsymbol{x}}$ and their respective cross-correlations.
2. The stochastic noise terms $\underline{\boldsymbol{w}}_k$ and $\underline{\boldsymbol{v}}_k$ are Gaussian, white and zero-mean centered, i.e. $\underline{\boldsymbol{w}}_k \sim \mathcal{N}(\underline{0}, \mathbf{C}_k^w)$ and $\underline{\boldsymbol{v}}_k \sim \mathcal{N}(\underline{0}, \mathbf{C}_k^z)$.
3. The state transition and measurement models $\underline{a}_k(\,\cdot\,)$ and $\underline{h}_k(\,\cdot\,)$ are linear functions of the state, the noise terms and the system input.

In this case, the first two moments, namely the mean and covariance, describe the probability density. The Kalman filter computes an estimate, $\underline{\hat{x}}_k$ that minimizes the mean squared error (MSE)

$$\mathrm{MSE}\left(\underline{\hat{x}}_k^{\mathrm{e}}\right) = \mathrm{E}\left[\left(\underline{\hat{x}}_k^{\mathrm{e}} - \underline{x}_k\right)^{\mathrm{T}}\left(\underline{\hat{x}}_k^{\mathrm{e}} - \underline{x}_k\right)\right] \tag{4.9}$$

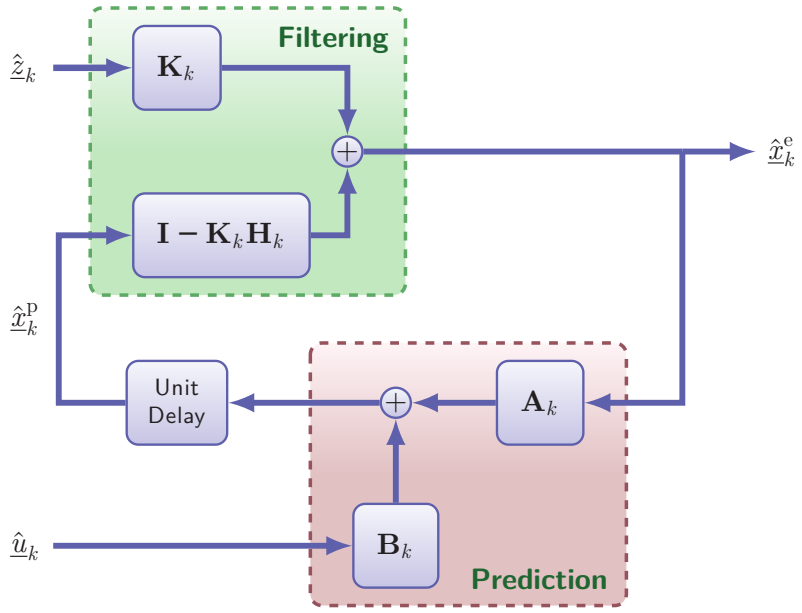

Figure 4.4: Kalman filter estimation consisting of prediction and filtering steps. Image source: Noack [62].

for the linear and Gaussian case. The Kalman filter recursively computes the covariance matrix

$$\mathbf{C}_k^{\mathrm{e}} = \mathrm{E}\left[\left(\hat{\underline{x}}_k^{\mathrm{e}} - \underline{x}_k\right)\left(\hat{\underline{x}}_k^{\mathrm{e}} - \underline{x}_k\right)^{\mathrm{T}}\right] \in \mathbb{R}^{n_x \times n_x}, \tag{4.10}$$

whose trace yields the MSE (4.9) [62].

Prediction

With a linear state transition model, $\mathbf{A}_k, \mathbf{B}_k$, the predicted estimate for an existing estimate $\hat{\underline{x}}_k^{\mathrm{e}}$ at time-step k is given by

$$\hat{\underline{x}}_{k+1}^{\mathrm{p}} = \mathbf{A}_k\hat{\underline{x}}_k^{\mathrm{e}} + \mathbf{B}_k\hat{\underline{u}}_k. \tag{4.11}$$

As the model noise $\underline{\boldsymbol{w}}_k$ is zero-mean centered, no further adjustment is needed for $\hat{\underline{x}}_{k+1}^{\mathrm{p}}$. The covariance matrix of the model noise $\mathbf{C}_k^w$ is used to handle this uncertainty in the predicted covariance

$$\mathbf{C}_{k+1}^{\mathrm{p}} = \mathbf{A}_k\mathbf{C}_k^{\mathrm{e}}\mathbf{A}_k^{\mathrm{T}} + \mathbf{B}_k\mathbf{C}_k^w\mathbf{B}_k^{\mathrm{T}} \tag{4.12}$$

Filtering

The predicted estimate $\underline{\hat{x}}_k^{\mathrm{p}}$ is linearly combined with the received measurement $\underline{\hat{z}}_k$ to obtain the optimal value that minimizes the trace of $\mathbf{C}_k^{\mathrm{e}}$. The Kalman gain factor describes how $\underline{\hat{x}}_k^{\mathrm{p}}$ and $\underline{\hat{z}}_k$ are to be weighted. The Kalman gain is expressed as

$$\mathbf{K}_k = \mathbf{C}_k^{\mathrm{p}} \mathbf{H}_k^{\mathrm{T}} \left(\mathbf{C}_k^z + \mathbf{H}_k \mathbf{C}_k^{\mathrm{p}} \mathbf{H}_k^{\mathrm{T}} \right)^{-1}, \tag{4.13}$$

where $\mathbf{H}_k$ is the linear measurement model that transforms the state space into the measurement space

$$\underline{z}_k = \mathbf{H}_k \underline{x}_k. \tag{4.14}$$

The Kalman gain $\mathbf{K}_k$ is used to compute the estimate

$$\begin{aligned} \underline{\hat{x}}_k^{\mathrm{e}} &= \underline{\hat{x}}_k^{\mathrm{p}} + \mathbf{K}_k \left(\underline{\hat{z}}_k - \mathbf{H}_k \underline{\hat{x}}_k^{\mathrm{p}} \right) \\ &= \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right) \underline{\hat{x}}_k^{\mathrm{p}} + \mathbf{K}_k \underline{\hat{z}}_k \end{aligned} \tag{4.15}$$

and the resulting covariance matrix

$$\begin{aligned} \mathbf{C}_k^{\mathrm{e}} &= \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right) \mathbf{C}_k^{\mathrm{p}} \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right)^{\mathrm{T}} + \mathbf{K}_k \mathbf{C}_k^z \mathbf{K}_k^{\mathrm{T}} \\ &= \mathbf{C}_k^{\mathrm{p}} - \mathbf{K}_k \mathbf{H}_k \mathbf{C}_k^{\mathrm{p}}. \end{aligned} \tag{4.16}$$

4.5 Dealing with Non-linearities

In reality, the previously listed conditions for using the linear Kalman filter are seldom met. Two widely used techniques, namely the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), are used for non-linear state transition and measurement models, $\underline{a}_k(\,\cdot\,)$ and $\underline{h}_k(\,\cdot\,)$. A Gaussian density passed through a non-linear function no longer remains Gaussian. Both of these techniques aim to approximate the resulting probability density with a Gaussian distribution. Due to this inexact approximation, the resulting filter is no longer optimal in reducing the MSE as in the linear case . Nevertheless, the possible application of the Kalman filter is broadened significantly because $\underline{a}_k(\,\cdot\,)$ and $\underline{h}_k(\,\cdot\,)$ are rarely linear. Conditions 1 and 2 from Section 4.4, namely zero-mean centered white noise terms and a Gaussian prior, are still assumed. For highly non-linear systems, other techniques such as the particle filter, are commonly in use. Due to time constraints and because non-linear distance equations are generally good natured [76], the focus remains on EKF and UKF.

4.5.1 Extended Kalman filter (EKF)

For using the Kalman filter for the non-linear case of estimating the position and velocity of a circumlunar trajectory, NASA implemented an extended version of the Kalman filter in 1962, which uses a first-order Taylor expansion to linearise the measurement and state transition models [80]. These are represented by the following Jacobians $\underline{x}_k \in \mathbb{R}^N$, $\underline{u}_k \in \mathbb{R}^L$ and $\underline{h}_k \in \mathbb{R}^M$.

$$\begin{aligned}\mathbf{A}_k &= \left.\frac{\partial \underline{a}_k(\,\cdot\,)}{\partial \underline{x}_k}\right|_{\underline{x}_k=\hat{\underline{x}}_k^{\mathrm{e}},\underline{u}_k=\hat{\underline{u}}_k} \\ &= \begin{bmatrix} \frac{\partial a_{k,1}}{\partial x_{k,1}} & \frac{\partial a_{k,1}}{\partial x_{k,2}} & \cdots & \frac{\partial a_{k,1}}{\partial x_{k,N}} \\ \frac{\partial a_{k,2}}{\partial x_{k,1}} & \frac{\partial a_{k,2}}{\partial x_{k,2}} & \cdots & \frac{\partial a_{k,2}}{\partial x_{k,N}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial a_{k,N}}{\partial x_{k,1}} & \frac{\partial a_{k,N}}{\partial x_{k,2}} & \cdots & \frac{\partial a_{k,N}}{\partial x_{k,N}} \end{bmatrix}_{\underline{x}_k=\hat{\underline{x}}_k^{\mathrm{e}},\underline{u}_k=\hat{\underline{u}}_k},\end{aligned} \tag{4.17}$$

$$\begin{aligned}\mathbf{B}_k &= \left.\frac{\partial \underline{a}_k(\,\cdot\,)}{\partial \underline{u}_k}\right|_{\underline{x}_k=\hat{\underline{x}}_k^{\mathrm{e}},\underline{u}_k=\hat{\underline{u}}_k} \\ &= \begin{bmatrix} \frac{\partial a_{k,1}}{\partial u_{k,1}} & \frac{\partial a_{k,1}}{\partial u_{k,2}} & \cdots & \frac{\partial a_{k,1}}{\partial u_{k,L}} \\ \frac{\partial a_{k,2}}{\partial u_{k,1}} & \frac{\partial a_{k,2}}{\partial u_{k,2}} & \cdots & \frac{\partial a_{k,2}}{\partial u_{k,L}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial a_{k,N}}{\partial u_{k,1}} & \frac{\partial a_{k,N}}{\partial u_{k,2}} & \cdots & \frac{\partial a_{k,N}}{\partial u_{k,L}} \end{bmatrix}_{\underline{x}_k=\hat{\underline{x}}_k^{\mathrm{e}},\underline{u}_k=\hat{\underline{u}}_k}\end{aligned} \tag{4.18}$$

and

$$\begin{aligned}\mathbf{H}_k &= \left.\frac{\partial \underline{h}_k(\,\cdot\,)}{\partial \underline{x}_k}\right|_{\underline{x}_k=\hat{\underline{x}}_{k+1}^{\mathrm{p}}} \\ &= \begin{bmatrix} \frac{\partial h_{k,1}}{\partial x_{k,1}} & \frac{\partial h_{k,1}}{\partial x_{k,2}} & \cdots & \frac{\partial h_{k,1}}{\partial x_{k,N}} \\ \frac{\partial h_{k,2}}{\partial x_{k,1}} & \frac{\partial h_{k,2}}{\partial x_{k,2}} & \cdots & \frac{\partial h_{k,2}}{\partial x_{k,N}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial h_{k,M}}{\partial x_{k,1}} & \frac{\partial h_{k,M}}{\partial x_{k,2}} & \cdots & \frac{\partial h_{k,M}}{\partial x_{k,N}} \end{bmatrix}_{\underline{x}_k=\hat{\underline{x}}_{k+1}^{\mathrm{p}}}\end{aligned} \tag{4.19}$$

The non-linear models are still used to compute the predicted estimate

$$\hat{\underline{x}}_{k+1}^{\mathrm{p}} = \underline{a}_k\left(\hat{\underline{x}}_k^{\mathrm{e}}, \hat{\underline{u}}_k\right) \tag{4.20}$$

and the filtered estimate

$$\hat{\underline{x}}_k^{\mathrm{e}} = \hat{\underline{x}}_k^{\mathrm{p}} + \mathbf{K}_k\left(\hat{\underline{z}}_k - \underline{h}_k\left(\hat{\underline{x}}_k^{\mathrm{p}}\right)\right). \tag{4.21}$$

The covariance update and Kalman gain equations, (4.12), (4.16) and (4.13), from the general Kalman filter can be used by substituting the Jacobians whenever necessary.

4.5.2 Unscented Kalman filter (UKF)

Julier and Uhlman proposed the UKF in 1997, which approximates the mean and covariance of the posterior by propagating *sigma points* through the non-linear state transition and measurement models [41].

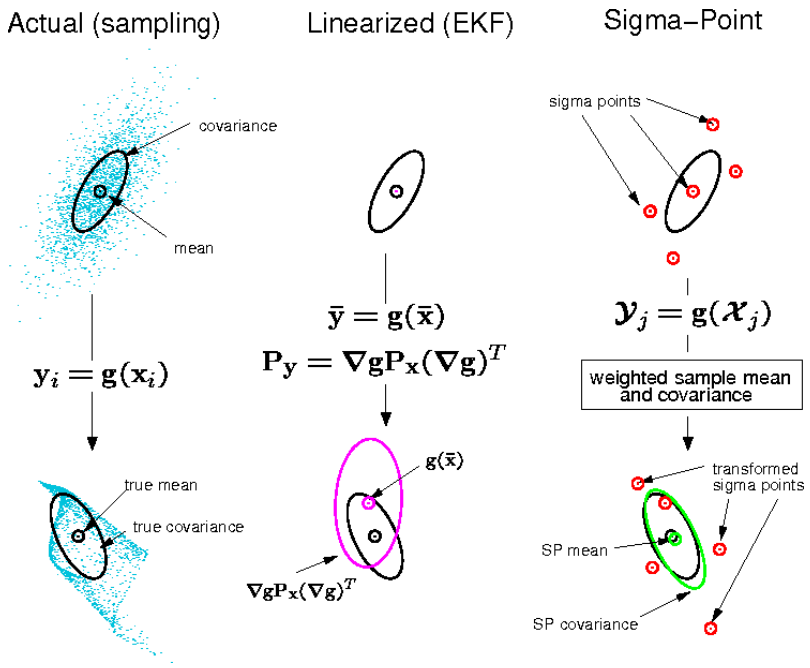

Figure 4.5: The UKF generates sigma points and propagates them through the non-linear models. As seen in this illustration, the non-linear state transition model results in a density that is no longer Gaussian. In this case, the Gaussian approximation by the UKF using the weighted sigma points is illustrated to better approximate the covariance of the Gaussian fit of the true density. Image source: Wan et al. [87].

For $\underline{x}_k \in \mathbb{R}^N$, it is common to generate $2N+1$ sigma points. Thus, the sigma points $\underline{\xi}_i, (i = 0, \ldots, 2N)$ for $\underline{\xi}_i \in \mathbb{R}^N$ are generated.. Though it is possible to vary this number, w.l.o.g, the UKF steps are explained for $2N+1$ sigma points

$$\underline{\xi}_i = \begin{cases} \underline{\hat{x}}_k^{\mathrm{e}}, i = 0 \\ \underline{\hat{x}}_k^{\mathrm{e}} + \left(\sqrt{(N+\lambda)\,\mathbf{C}_k^{\mathrm{e}}}\right)_i, \; i = 1, \ldots, N \\ \underline{\hat{x}}_k^{\mathrm{e}} - \left(\sqrt{(N+\lambda)\,\mathbf{C}_k^{\mathrm{e}}}\right)_i, \; i = N+1, \ldots, 2N. \end{cases} \tag{4.22}$$

$\left(\sqrt{(N+\lambda)\,\mathbf{C}_k^{\mathrm{e}}}\right)_i$ merely denotes the i-th column of the square root decomposition of the covariance matrix. This is provided by selecting the lower triangular matrix $\sqrt{\mathbf{C}_k^{\mathrm{e}}}$ of the Cholesky decomposition $\mathbf{C}_k^{\mathrm{e}} = \sqrt{\mathbf{C}_k^{\mathrm{e}}}\sqrt{\mathbf{C}_k^{\mathrm{e}}}^{\mathrm{T}}$. λ is a scaling factor

$$\lambda = \alpha^2\,(N+\kappa) - N, \tag{4.23}$$

where $\kappa \geq 0$ and $\alpha \in (0, 1]$ influence how far the sigma points are away from the mean. The weighting factors for the sigma points, w_m^i and w_c^i, are calculated as

$$\begin{aligned} w_0^m &= \frac{\lambda}{n+\lambda} \\ w_0^m &= w_0^m + \left(1 - \alpha^2 + \beta\right) \\ w_i^m = w_i^c &= \frac{1}{2(n+\lambda)} \quad \text{for } i = 1, \ldots, 2n, \end{aligned} \tag{4.24}$$

where β is another factor that is set according to the distribution of $\underline{\boldsymbol{x}}$. For a Gaussian distribution, a value of $\beta = 2$ is optimal.

Prediction

The set of sigma points $\{\underline{\xi}_0, \ldots, \underline{\xi}_{2N}\}$ are propagated through the non-linear state transition model $\underline{a}_k(\,\cdot\,)$

$$\underline{\xi}_i^{\mathrm{p}} = \underline{a}_k(\underline{\xi}_i, \underline{\hat{u}}_k),\ i = 0, \ldots, 2N. \tag{4.25}$$

Using the weights allows us to compute the predicted state vector and covariance

$$\underline{\hat{x}}_{k+1}^{\mathrm{p}} = \sum_{i=0}^{2N} w_i^{(m)} \underline{\xi}_i^{\mathrm{p}} \tag{4.26}$$

$$\mathbf{C}_{k+1}^{\mathrm{p}} = \sum_{i=0}^{2N} w_i^{(c)} \left(\underline{\xi}_i^{\mathrm{p}} - \underline{\hat{x}}_{k+1}^{\mathrm{p}}\right)\left(\underline{\xi}_i^{\mathrm{p}} - \underline{\hat{x}}_{k+1}^{\mathrm{p}}\right)^{\mathrm{T}} + \mathbf{C}_k^{w} \tag{4.27}$$

Filtering

The next time step k, at which the measurement $\underline{\hat{z}}_k$ arrives, is now considered. The predicted sigma points are transformed to the measurement space using the measurement model

$$\underline{\zeta}_i = \underline{h}_k(\underline{\xi}_i^{\mathrm{p}}),\ i = 0, \ldots, 2N, \tag{4.28}$$

which can be used to calculated the predicted value of the measurement and its respective covariance

$$\underline{\hat{z}}_k^{\mathrm{p}} = \sum_{i=0}^{2N} w_i^{(m)} \underline{\zeta}_i \tag{4.29}$$

$$\mathbf{C}_k^{\mathrm{S}} = \sum_{i=0}^{2N} w_i^{(c)} \left(\underline{\zeta}_i - \underline{\hat{z}}_k^{\mathrm{p}}\right) \left(\underline{\zeta}_i - \underline{\hat{z}}_k^{\mathrm{p}}\right)^{\mathrm{T}} + \mathbf{C}_k^z. \tag{4.30}$$

The cross-covariance between the transformed sigma points $\underline{\zeta}_i$ and the untransformed sigma points $\underline{\xi}_i^p$

$$\mathbf{C}_k^{xz} = \sum_{i=0}^{2N} w_i^{(c)} \left(\underline{\zeta}_i - \underline{\hat{z}}_k^{\mathrm{p}}\right) \left(\underline{\xi}_i^{\mathrm{p}} - \underline{\hat{x}}_k^{\mathrm{p}}\right)^{\mathrm{T}} \tag{4.31}$$

is also required.

The Kalman gain is calculated with these two covariance matrices

$$\mathbf{K}_k = \mathbf{C}_k^{xz} \left(\mathbf{C}_k^{\mathrm{S}}\right)^{-1}, \tag{4.32}$$

which is then used to update the state vector and its covariance using the received measurement $\underline{\hat{z}}_k$ by applying the equations

$$\underline{\hat{x}}_k^{\mathrm{e}} = \underline{\hat{x}}_k^{\mathrm{p}} + \mathbf{K}_k \left(\underline{\hat{z}}_k - \underline{\hat{z}}_k^{\mathrm{p}}\right) \text{ and} \tag{4.33}$$

$$\mathbf{C}_k^{\mathrm{e}} = \mathbf{C}_k^{\mathrm{p}} - \mathbf{K}_k \mathbf{C}_k^{\mathrm{S}} \mathbf{K}_k^{\mathrm{T}}. \tag{4.34}$$

4.6 Summary

In Section 4.2, the reasons for using state estimation techniques are listed. The general Bayesian estimator was explained in Section 4.3. Though it is the optimal state estimator, it is clear that a real-world usage is not possible. As described in Section 4.4, the Kalman filter is the optimal state estimator if certain conditions are met. For non-linear state transition and measurement models, two popular variants of the Kalman filter, namely EKF and UKF, are also explained. These techniques play a vital role in designing centralized and distributive state estimation techniques for goal of WSN localization. Before the application of these state estimators can be shown, certain important fundamentals have to be clarified for handling multi-sensor measurements in a WSN.