Advanced Autonomous Systems

Bayesian Estimation: The Kalman Filter

As we mentioned before, one remarkable aspect of the Bayesian approach is its ability to fuse different sources of information. However, its implementation can be difficult in terms of processing cost and memory usage. For instance, for a given observation (usually due to a measurement), we need to synthesize the likelihood function and then multiply it to the prior PDF about the estimated state, to finally obtain the updated PDF of the state. All the factors involved in this multiplication are functions of the variable **X**. The state vector **X** can be, depending on the problem, a 1D, or 2D or 3D vector, or, can even be a high dimensional variable. For low dimensional cases (e.g. 1D or 2D) it is feasible to numerically manipulate the arbitrary PDFs; however, for higher dimensions it is usually difficult and expensive to do it in that way, in particular if the processing has to be performed in real-time.
Some simplifications for this problem can help, to make the solution of this problem dramatically less expensive, i.e. making it feasible of being performed in real-time.

As we discussed before, if a RV **X** is described by a Gaussian PDF, any linear transformation of **X** will result in a new RV, which will be Gaussian as well. In addition, if Gaussian noise is added to a Gaussian RV the resulting RV is Gaussian as well.
If an observation is given, of the type $0 = h(\mathbf{X}) + \boldsymbol{\eta}$, in which the function $h(\mathbf{X})$ is linear, and the noise $\boldsymbol{\eta}$ is Gaussian and white (GWN), then the likelihood function associated to this observation will be of the Gaussian type.
If we have a Gaussian PRIOR PDF for describing the variable **X**, and we apply a Bayesian update associated to a Gaussian likelihood function, then the resulting POSTERIOR PDF will be of the Gaussian type as well. This is because the multiplication of two functions of the Gaussian type gives a resulting function which is Gaussian function.
In addition to these interesting properties, when operating with Gaussian PDFs and linear transformations there is an additional and very convenient aspect: We do not need to deal with the PDFs themselves but with their parameters, i.e. their expected values and their covariance matrixes. Those entities are just numbers (vectors and matrixes); which is also very convenient because working with vectors and matrixes is definitely easier than dealing with multivariate functions.

Calculation of the PDF for a transformed variable:

Consider the following Gaussian RVs

$$\mathbf{X} \in \mathbb{R}^N, \mathbf{X} \sim N\left(\mathbf{X}; \hat{\mathbf{X}}, \mathbf{P}_{\mathbf{X}}\right)$$

$$\boldsymbol{\zeta} \in \mathbb{R}^M, \quad \boldsymbol{\zeta} \sim N\left(\boldsymbol{\zeta}; \mathbf{0}, \mathbf{P}_{\boldsymbol{\zeta}}\right)$$

In addition, the RVs are independent, i.e. they cross-covariance is a zero matrix (*),

$$E\left\{\left(\mathbf{X} - \hat{\mathbf{X}}\right) \cdot \left(\boldsymbol{\zeta} - \hat{\boldsymbol{\zeta}}\right)^T\right\} = \overline{\overline{\mathbf{0}}}$$

(*) It is a necessary and sufficient condition for saying that both Gaussian RVs are independent; however, it is not a general condition for other PDF classes. We briefly discuss about the interpretation of it, in class, considering a joint PDF which jointly describes both RVs.

If we have a linear transformation of the form

$$\mathbf{Z} = \mathbf{A} \cdot \mathbf{X} + \boldsymbol{\zeta}, \qquad \mathbf{A} \in \mathbb{R}^{M \cdot N}, \qquad \text{(E1)}$$

Then the PDF of the resulting variable $\mathbf{Z}$ is a Gaussian function, $\mathbf{Z} \sim N\left(\mathbf{Z}; \hat{\mathbf{Z}}, \mathbf{P}_{\mathbf{Z}}\right)$, and its expected value and covariance matrix are simply given by:

$$\begin{aligned} \hat{\mathbf{Z}} &= \mathbf{A} \cdot \hat{\mathbf{X}} + \bar{\mathbf{0}} \\ \mathbf{P}_{\mathbf{Z}} &= \mathbf{A} \cdot \mathbf{P}_{\mathbf{X}} \cdot \mathbf{A}^T + \mathbf{P}_{\zeta} \end{aligned} \qquad \text{(E2)}$$

(Note that the size of the vectors and matrixes involved in the calculations are consistent, i.e.

$$\hat{\mathbf{Z}} \in \mathbb{R}^M, \quad \mathbf{P}_{\mathbf{Z}} \in \mathbb{R}^{M \cdot M}, \quad \hat{\mathbf{X}} \in \mathbb{R}^N, \quad \mathbf{P}_{\mathbf{X}} \in \mathbb{R}^{N \cdot N}, \quad \hat{\boldsymbol{\zeta}} = 0 \in \mathbb{R}^M, \quad \mathbf{P}_{\zeta} \in \mathbb{R}^{M \cdot M}, \quad \mathbf{A} \in \mathbb{R}^{M \cdot N})$$

Prediction of a Random Variable based on a Process Model

Previously, we considered a particular case of a linear combination of RVs, because it can explain the case of a linear Process Model. We have a system whose model is linear, and whose uncertainty is represented by a zero mean Gaussian R.V. Our case is the following one,

$$\mathbf{X}(k+1) = \mathbf{A} \cdot \mathbf{X}(k) + \boldsymbol{\zeta}(k) \qquad \text{(E3)}$$

The variable $\mathbf{X}(k+1)$ is just the case of a variable which is function of two other variables (i.e. $\mathbf{X}(k), \boldsymbol{\zeta}(k)$), as it is expressed in E1. Consequently, we can apply (E2) in order to obtain the statistical description of our belief about $\mathbf{X}(k+1)$.

This type of operation is usually called ***"prediction step"*** because the estimate of $\mathbf{X}(k+1)$ is purely based on an estimate of $\mathbf{X}(k)$ and a process model. The estimate of the predicted $\mathbf{X}(k+1)$ is expressed as a PDF and this PDF is the result of considering the following information:

1) A PDF which describes the knowledge about the variable $\mathbf{X}(k)$.
2) A PDF which describes the uncertainty/noise ($\boldsymbol{\zeta}(k)$), which pollutes the assumed process model.
3) The assumed process model itself, (E3)

For (1) and (2), the expected values and covariance matrixes, associated to the PDFs, are needed.
Part (3) is provided by the mathematical description of the process model, (E3)

The usual notation for expressing the result of a prediction step is through the couple of parameters $\hat{\mathbf{X}}(k+1|k)$, $\mathbf{P}(k+1|k)$, i.e. the expected value and covariance matrix of the PDF. This notation means that those expected value and covariance matrix do correspond to the estimates about $\mathbf{X}(k+1)$ just based on the estimates of $\mathbf{X}(k)$, and on the assumed process model. No fresh observations (which may be available at time *k+1*) are considered for obtaining these estimates about $\mathbf{X}(k+1)$. If information from the observations that are available at time *k+1* were also considered for obtaining the estimates, then the expected value and covariance matrix would be expressed as $\hat{\mathbf{X}}(k+1|k+1)$ and $\mathbf{P}(k+1|k+1)$, respectively.

We are using the notation $\hat{\mathbf{X}}(i|j)$, $\mathbf{P}(i|j)$, to express the expected value and covariance matrix of the estimates of the variable $\mathbf{X}(i)$, based on observations provided from time *k=0* up to time *k=j* .

The prediction steps can be performed by applying the following matrix operations:

$$\begin{aligned}\hat{\mathbf{X}}(k+1|k) &= \mathbf{A}\cdot\hat{\mathbf{X}}(k|k)\\ \mathbf{P}(k+1|k) &= \mathbf{A}\cdot\mathbf{P}(k|k)\cdot\mathbf{A}^T+\mathbf{Q}(k)\end{aligned} \qquad \text{(E4)}$$

where the matrix $\mathbf{Q}(k)$ is the covariance of the process noise $\boldsymbol{\zeta}(k)$.
Note that the more general case, which considers inputs, $\mathbf{X}(k+1)=\mathbf{A}\cdot\mathbf{X}(k)+\mathbf{B}\cdot\mathbf{u}(k)+\boldsymbol{\zeta}(k)$, is treated in a similar way. If the input (variable $\mathbf{u}(k)$) is perfectly known then it is treated as an additive deterministic component of the process model expression, only affecting the expected value of $\mathbf{X}(k+1)$, i.e. $\hat{\mathbf{X}}(k+1|k)=\mathbf{A}\cdot\hat{\mathbf{X}}(k|k)+\mathbf{B}\cdot\mathbf{u}(k)$. In cases in which uncertainty is present in our knowledge about $\mathbf{u}(k)$, this additional source of uncertainty will be treated as a RV as well, and it will contribute to the assumed noise in the process model. How to do this additional treatment will be discussed later this lecture.

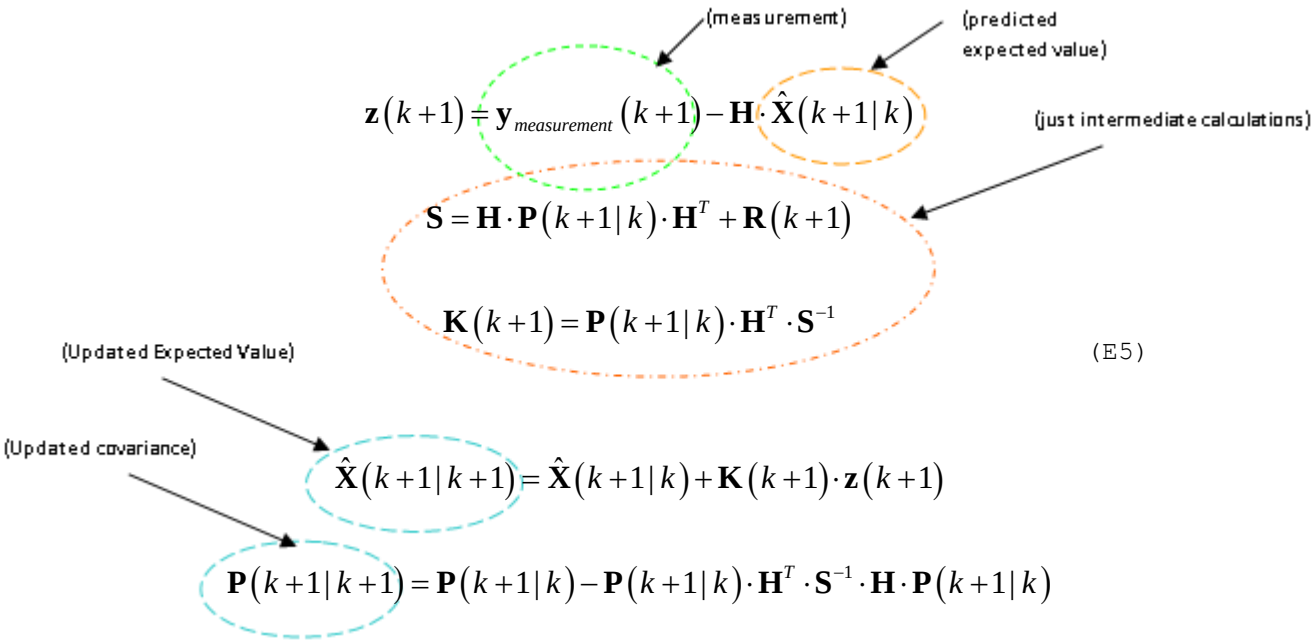
Update of a Random Variable due to an Observation

Suppose that at time *k+1* some observation, which involves the state vector $\mathbf{X}(k+1)$, does happen. For instance, suppose a measurement of certain output of the system, such as the following one,

$$\mathbf{y}(k+1)=\mathbf{H}\cdot\mathbf{X}(k+1)+\boldsymbol{\eta}(k+1)$$

$$\mathbf{y}(k+1)\in\mathbb{R}^L,\quad \boldsymbol{\eta}(k+1)\in\mathbb{R}^L$$
$$\mathbf{H}\in\mathbb{R}^{L\times N}$$

where the uncertainty component $\boldsymbol{\eta}(k+1)$ is assumed to be zero mean Gaussian noise (more exactly it is assumed to be Gaussian *white noise*, a concept which was explained in "Lecture 0"). This noise is known to have certain covariance, whose value is indicated by the matrix $\mathbf{R}(k+1)$.
This observation introduces a constraint on our belief about $\mathbf{X}(k+1)$; this constraint can be statistically described by a likelihood function. This likelihood function has a Gaussian shape, and, after multiplying the PRIOR (the previously predicted PDF about $\mathbf{X}(k+1)$), it will produce a new Gaussian PDF, whose mean and covariance take the values which are expressed in (E5). The resulting mean is called $\hat{\mathbf{X}}(k+1|k+1)$, which represents the expected value of $\mathbf{X}(k+1)$ based on observations obtained up to time *k+1*, i.e. also including the one provided by the measurement (or measurements) $\mathbf{y}(k+1)$. Similarly, the updated covariance is denoted as $\mathbf{P}(k+1|k+1)$.
A way for performing the update is expressed by the following sequence of operations:



This sequence of operations is performed for obtaining the POSTERIOR PDF, which is fully represented by the expected value $\hat{\mathbf{X}}(k+1|k+1)$ and covariance $\mathbf{P}(k+1|k+1)$; which are the parameters we want to know. The rest of the matrixes and vectors, which are involved in (E5), are just intermediate calculations, needed for calculating what we want. The calculations expressed in (E5) are a way for obtaining the resulting Gaussian PDF, which is the result of multiplying the PRIOR PDF and the likelihood function (both Gaussian functions as well). This is what we discussed in previous lectures about the Bayesian approach, for processing observations!

(btw: We do not need to memorize these expressions in E5; however, it must be clear why we are applying them: it is a way of implementing the Bayes' rule!)

The resulting value $\mathbf{z}(k+1)$ is called the innovation; it is the discrepancy (residual) between the expected output (expected measurement) and the actual measurement. (Note: usually the real value of the outputs is neither given by the measurement nor the expected one; question: Why are we saying this?)
The matrix $\mathbf{S}$ is an estimate of the covariance of $\mathbf{z}(k+1)$. See the terms that participate in its calculation: $\mathbf{P}(k+1|k)$ and $\mathbf{R}(k+1)$. The matrix $\mathbf{R}(i)$ represents the covariance of the measurement noise at time ***i***. (We will discuss, in the lecture, about values for typical cases, e.g. for the LIDAR we use in the projects.)
The resulting *gain* matrix $\mathbf{K}$ defines how to update the expected value of $\mathbf{X}(k+1)$ and its covariance.
The resulting updated covariance is shown as $\mathbf{P}(k+1|k+1)$, which means it is the covariance of the estimates of $\mathbf{X}(k+1)$, which are based on the information of all the available observations up to time *k+1* (including the ones at that time).

Through this approach, we have a very efficient recursive method to keep the estimates of a dynamic system whose dynamics is modeled by a linear system, whose uncertainty is assumed to be GWN.
We have also some measurements whose relationship with the internal state of the system is linear, and whose uncertainty is modeled as a Gaussian RV.
This means that, in order to maintain estimates about **X** at all times, we just need to keep recursively applying operations such as in (E4) and (E5). Those usually define a sequence of steps, like the following one:

…→Prediction → observation → Prediction → observation → Prediction →observation →…,

i.e. we predict $\mathbf{X}(i)$ based on measurements till time $(i-1)$. We apply the update of it, based on measurements of the output of the system at time ***i***; then we predict $\mathbf{X}(i+1)$ based on our belief about $\mathbf{X}(i)$, and the assumed process model; then we update its current belief by considering the measurements at time $(i+1)$, a so on…

$$(\ ..\begin{Bmatrix}\hat{\mathbf{X}}(i|i-1)\\ \mathbf{P}(i|i-1)\end{Bmatrix}\xrightarrow[update]{\mathbf{y}(i)=h(\mathbf{x}(i))}\begin{Bmatrix}\hat{\mathbf{X}}(i|i)\\ \mathbf{P}(i|i)\end{Bmatrix}\xrightarrow[prediction]{\mathbf{x}(i+1)=F(\mathbf{x}(i),\mathbf{u}(i))}\begin{Bmatrix}\hat{\mathbf{X}}(i+1|i)\\ \mathbf{P}(i+1|i)\end{Bmatrix}\xrightarrow[update]{\mathbf{y}(i+1)=h(\mathbf{x}(i+1))}\begin{Bmatrix}\hat{\mathbf{X}}(i+1|i+1)\\ \mathbf{P}(i+1|i+1)\end{Bmatrix}\rightarrow..$$

For instance, we start the estimation process at time *0* (*k=0*). We have an initial estimate of **X**(0). Its statistics are defined by its expected value and covariance matrix, $\hat{\mathbf{X}}(0|0),\mathbf{P}(0|0)$**.**

We "move" to time ***k=1*** by performing a prediction step for obtaining $\hat{\mathbf{X}}(1|0),\mathbf{P}(1|0)$ (as described in (E4)), that is: these values describe $\mathbf{X}(1)$ but based on information provided before time ***k=1***.

Suppose that at time *k=1* there is an observation (a vector of measurements of the system's outputs, $\mathbf{y}(1)=\mathbf{H}(1)\cdot\mathbf{X}(1)$), consequently we are able to update the estimates for obtaining $\hat{\mathbf{X}}(1|1),\mathbf{P}(1|1)$. The time evolves to *k=2*, and now we need to know the current value of **X** at that time, i.e. we need an estimate about $\mathbf{X}(2)$. We can initially perform a prediction step as in (E4), for producing $\hat{\mathbf{X}}(2|1),\mathbf{P}(2|1)$. If there is an observation available at time *k=2* (such as certain observation $\mathbf{y}(2)=\mathbf{H}(2)\cdot\mathbf{X}(2)$) then we proceed by performing an update operation for generating $\hat{\mathbf{X}}(2|2),\mathbf{P}(2|2)$. Note that we expressed this linear operation by using the matrix $\mathbf{H}(2)$; this is because the observation functions can be different ones at different times (e.g. due to a time varying system, or due to using different sensors having output equations, etc.)

Similar situation may happen with the process model, it can be time-varying (although still linear). Even the statistical properties of the noise can change with the time, for instance the covariance of $\boldsymbol{\eta}(k)$ may be different (but known) at different times *k*. This is why we expressed the noises' covariance matrixes as functions of the time, $\mathbf{Q}(k)$ and $\mathbf{R}(k)$.

Note that what we are now describing, in this section, is the Bayesian Estimation process we had discussed for the general case; however, we are now considering the Gaussian case (which simplifies our calculation, enormously!).

Asynchronous Operation

There is nothing wrong in having consecutive prediction steps without having updates (it means there were no observations in that period of time). It is also possible to have multiple simultaneous observations. If these observations are independent (their noises have no dependency between them) then they can be processed individually, in a sequential fashion (to be discussed in the lecture).

An estimation process, which performs a sequence like the one we discussed in the previous sections, is called a KALMAN FILTER.

Example:

In this example we see how to obtain the matrixes related to the process and observation models.

Suppose a system composed by an electric motor, a pulley and a load. Suppose we control the input current, which feeds the motor. E.g. by using a DAC we control the voltage which is the input of an amplifier

(voltage/current) which feeds the current to the motor. Suppose we sample the angular position every 4ms (through an encoder) and that there are some errors in the measurement (encoder quantization, etc.).
We consider that the differential equation for describing the evolution of the angular position is $\ddot{\theta}(t) = -a \cdot \dot{\theta}(t) + b \cdot u(t)$

Where $\mathbf{u}(t)$ is the input (the torque, proportional to the voltage from the computer's DAC), the coefficients ***a*** and ***b*** in the model are defined by the friction, moment of inertia, voltage/torque gain, etc. We know the values (but not perfectly) of the coefficients and we know the value of $\mathbf{u}(t)$ at each sample time (we produce that signal). However, we do not know the values of the angular position and speed. We measure, sporadically, the angular position and these measurements are polluted by errors.
We consider the following state space representation and process model,

$$\mathbf{X}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \theta(t) \\ \dot{\theta}(t) \end{bmatrix}$$

$$\Downarrow$$

$$\dot{\mathbf{X}}(t) = \begin{bmatrix} x_2(t) \\ -a \cdot x_2(t) + b \cdot u(t) \end{bmatrix} \qquad \text{(E6)}$$

If we have a discrete time control signal then it is possible to obtain the discrete time version of (E6) (the *Transition Matrix*, from Linear Control Theory)

$$\mathbf{X}(t+\tau) = \mathbf{A} \cdot \mathbf{X}(t) + \mathbf{B} \cdot \mathbf{u}(t)$$

We can avoid the notation that explicitly uses the time by using one which is expressed as function of the discrete time ***k*** (i.e. in steps):

$$\mathbf{X}(k+1) = \mathbf{A} \cdot \mathbf{X}(k) + \mathbf{B} \cdot \mathbf{u}(k)$$

This model belongs to the family described in E4, except that we have an extra component, i.e. due to the inputs.

$$\mathbf{X}(k+1) = \mathbf{A} \cdot \mathbf{X}(k) + \mathbf{B} \cdot \mathbf{u}(k) + \boldsymbol{\zeta}(k)$$

The term due to the inputs is treated as a constant (i.e. it is a deterministic term). If there is some uncertainty about the values of the inputs, then these can be described in the term that represents the uncertainty.
Details about this procedure are presented later, during the lecture.

At certain sample times we may perform a measurement. Suppose the sensor does directly measure the position of the load. In this particular case the observation function would be the following one:

$$\mathbf{Y}(k+1) = \mathbf{H} \cdot \mathbf{X}(k+1), \qquad \mathbf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

This function is independent of the angular speed (x_2), however it provides information to estimate the speed as well. How is this possible? The answer is that the process model also has a role in the observability of this system (remember *observability* from Control Theory).

The process model modifies the joint PDF (of x_1, x_2) in a way that makes the RVs to be correlated (they are not independent). Then, by getting information about x_1 we also introduce knowledge about x_2. The estimator can then estimate both variables, x_1 and x_2 (position and speed), in a simultaneous fashion.

EXTENDED KALMAN FILTER

In the previous section we discussed about the Kalman Filter, i.e. an estimation process in which all the involved PDFs are Gaussian, and the models (for process and observations) do involve linear operations (note: a bit more than linear: Affine operations, $\mathbf{A}\cdot\mathbf{X}+\mathbf{B}$).
The systems we need to treat are usually NON-LINEAR, and the PDFs are NON-GAUSSIAN; however, if the models can be approximated (locally) by linear models, and the PDFs are unimodal and can also be approximated by Gaussian PDFs, then it is possible to apply a modified version of the KF method. A known approach to do it is the Extended KF (EKF). The EKF is a well-known variant of the KF, however there are other variants of the KF which are also intended to solve the same class of problems (we do not discuss about those alternative approaches, in this course.)
As we discussed in a previous lecture, those non-linear functions can be approximated by first order Taylor expansions. The trick in the EKF approach is that it performs those linear approximations at the most probable values of the estimated state, i.e. at the expected values. Then, when it performs an update due to a non-linear observation like $\mathbf{Y}(k+1)=h(\mathbf{X}(k+1))$, it applies the original KF update, but slightly modified. First, it obtains the **H** matrix by evaluating the Jacobian matrix of the observation function $h(\mathbf{X})$,

$$\mathbf{H}=\left.\frac{\partial h(\mathbf{X})}{\partial \mathbf{X}}\right|_{\mathbf{X}=\hat{\mathbf{X}}(k+1)}$$

i.e. it uses its best knowledge (about $\mathbf{X}(k+1)$) to choose the "point of operation" where to linearize (usually it will be at $\hat{\mathbf{X}}(k+1|k)$).

The expected measurement, $\hat{\mathbf{y}}(k+1)$, is obtained by directly evaluating $h(\mathbf{X})$ at the current expected value of $\mathbf{X}(k+1)$, i.e. $\hat{\mathbf{y}}(k+1)=h(\hat{\mathbf{X}}(k+1|k))$.

The innovation, as in the KF case, is the difference between the actual measurement and the expected measurement,

$$\mathbf{z}(k+1)=\mathbf{y}_{measurement}(k+1)-h(\hat{\mathbf{X}}(k+1|k))$$

This difference between the expected measurement and the real measurement will be used to update the knowledge about $\mathbf{X}(k+1)$, i.e. to produce the expected value $\hat{\mathbf{X}}(k+1|k+1)$, as in the standard KF (eq. E6).

A linearization process is also used in the prediction steps. If the process model is expressed as $\mathbf{X}(k+1)=f(\mathbf{X}(k),\mathbf{u}(k))+\boldsymbol{\zeta}(k)$, then we predict the expected value and covariance matrix by applying the following operations. For obtaining the predicted expected value we simply apply the process model equation,

$$\hat{\mathbf{X}}(k+1|k)=f(\hat{\mathbf{X}}(k|k),\mathbf{u}(k));$$

for evaluating the covariance matrix, we obtain the Jacobian matrix of the process model, evaluated at the current expected value,

$$\mathbf{F} = \left.\frac{\partial f\left(\mathbf{X},\mathbf{u}\right)}{\partial \mathbf{X}}\right|_{\mathbf{X}=\hat{\mathbf{X}}(k|k),\mathbf{u}=\mathbf{u}(k)}$$

and then we use it for evaluating the covariance matrix, by performing

$$\mathbf{P}\left(k+1|k\right) = \mathbf{F}\cdot\mathbf{P}\left(k|k\right)\cdot\mathbf{F}^{T} + \mathbf{Q}\left(k\right)$$

where the meaning of the items $\hat{\mathbf{X}}\left(k|k\right), \hat{\mathbf{X}}\left(k+1|k\right), \mathbf{P}\left(k|k\right), \mathbf{P}\left(k+1|k\right), \mathbf{Q}\left(k\right), \mathbf{R}\left(k\right)$ are identical to those of the standard KF estimator.

As the estimation process (KF or EKF) is kept recursively running in a periodic fashion, it will keep producing estimates of the system's states. We can use those estimates for our specific purposes, e.g. for controlling the system via state feedback.

Depending on the availability at each time ***k***, we may use $\hat{\mathbf{X}}\left(k|k\right)$ or $\hat{\mathbf{X}}\left(k|k-1\right)$. Naturally we would prefer to use $\hat{\mathbf{X}}\left(k|k\right)$, because it is the most informative version of the estimates of $\mathbf{X}\left(k\right)$.

We remark the fact that, in addition to the expected value, the covariance matrix $\mathbf{P}$ gives us an idea about the quality of the estimates, so when we use $\hat{\mathbf{X}}$, we know how accurate can be.

To see in class: A Matlab implementation of an example will be presented in the lecture ("Range and Bearing Localization").

Additional Considerations

We can also treat some cases where certain sources of uncertainty are not originally expressed as additive terms, such in the following general expression

$$\mathbf{y} = h\left(\hat{\mathbf{X}}\left(k+1|k\right), \boldsymbol{\eta}\left(k+1\right)\right)$$

In some cases, these contributions can be also approximated by linearization, as follows,

$$h\left(\mathbf{X},\boldsymbol{\eta}\right) \cong h\left(\mathbf{X},\hat{\boldsymbol{\eta}}\right) + \mathbf{B}\cdot\left(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}\right), \qquad \mathbf{B} = \left.\frac{\partial h}{\partial \boldsymbol{\eta}}\right|_{\mathbf{X}=\hat{\mathbf{X}},\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}}$$

For the cases where $\boldsymbol{\eta}$ is a zero mean Gaussian white noise, it becomes as follows,

$$h\left(\mathbf{X},\boldsymbol{\eta}\right) \cong h\left(\mathbf{X},\bar{\mathbf{0}}\right) + \mathbf{B}\cdot\boldsymbol{\eta}, \qquad \mathbf{B} = \left.\frac{\partial h}{\partial \boldsymbol{\eta}}\right|_{\mathbf{X}=\hat{\mathbf{X}},\boldsymbol{\eta}=\bar{\mathbf{0}}}$$

Consequently, in this approximation, the uncertainty is properly transformed to be simply treated as an additive source of noise.

Similar reasoning is applied for process models that involve non-linear expressions about certain noise. A usual case is when our knowledge about the system's inputs is not perfect,

$$\mathbf{X}(k+1) = f\left(\mathbf{X}(k), \mathbf{u}(k) + \boldsymbol{\xi}_{\mathbf{u}}(k)\right)$$

Where our assumed system's input, $\mathbf{u}(k)$, is polluted by uncertainty. In this case, the real input to the system is $\mathbf{u}^*(k)$, however only a polluted version of it, , $\mathbf{u}(k) = \mathbf{u}^*(k) - \boldsymbol{\xi}_{\mathbf{u}}(k)$, is available to us. For that reason, we replace the real (but unknown) input $\mathbf{u}^*(k)$ by $\mathbf{u}(k) + \boldsymbol{\xi}_{\mathbf{u}}(k)$.
An approximation for this function is the following one:

$$\mathbf{X}(k+1) = f\left(\mathbf{X}(k), \mathbf{u}(k)\right) + \mathbf{B} \cdot \boldsymbol{\xi}_{\mathbf{u}}(k), \qquad \mathbf{B} = \left.\frac{\partial f}{\partial \mathbf{u}}\right|_{\mathbf{X}=\hat{\mathbf{X}}(k), \mathbf{u}=\mathbf{u}(k)}$$

In which the term $\mathbf{u}(k)$ means the inputs we assume are applied to the system. This may be because we measure them or because we generate or apply them, but those nominal values are later polluted with some noise before are really applied to the system (e.g. quantization error in DAC, distortion in amplifiers, etc.)

Usually, if we have a source of uncertainty involved in any of the models (process or observation), we can treat it by proper linearization (considering the limitations of the approximation).

Interesting examples, not related to Robotics, to see in class.

1) Estimating states of a pendulum, for control, just measuring a noisy low frequency encoder.
2) Estimating states of system of interconnected tanks.

References:

Grewal. M.S. *"Kalman Filtering. Theory and Practice"*. Prentice Hall. 1993
Maybeck, P. S. *"Stochastic Models. Estimation and Control"* Vol I, Academic Press, New York, 1979
Jazwinski, A. H., *"Stochastic processes and filtering theory"*, Academic Press, 1970.