Bayesian Estimation: The EKF

Prediction Step

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

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

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

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

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

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

UPDATE Step

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

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

$$\mathbf{S} = \mathbf{H}\cdot\mathbf{P}(k+1|k)\cdot\mathbf{H}^T + \mathbf{R}(k+1)$$

$$\mathbf{K}(k+1) = \mathbf{P}(k+1|k)\cdot\mathbf{H}^T\cdot\mathbf{S}^{-1} \qquad \text{(E1)}$$

$$\hat{\mathbf{X}}(k+1|k+1) = \hat{\mathbf{X}}(k+1|k) + \mathbf{K}(k+1)\cdot\mathbf{z}(k+1)$$

$$\mathbf{P}(k+1|k+1) = \mathbf{P}(k+1|k) - \mathbf{P}(k+1|k)\cdot\mathbf{H}^T\cdot\mathbf{S}^{-1}\cdot\mathbf{H}\cdot\mathbf{P}(k+1|k)$$

…→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…

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

For instance, we start the estimation process at time *0* (*k=0*). We have an initial estimate about **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 $\hat{\mathbf{X}}(1\,|\,0), \mathbf{P}(1\,|\,0)$, that is: these values describe $\mathbf{X}(1)$ but based on observations provided before time ***k=1***.

Suppose that at time *k=1* there is an observation (a vector of measurements of the system's outputs, $\mathbf{y}(1) = h(\mathbf{X}(1))$), 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, 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 $\mathbf{y}(2) = h(\mathbf{X}(2))$) then we proceed by performing an update operation for generating $\hat{\mathbf{X}}(2\,|\,2), \mathbf{P}(2\,|\,2)$.

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.

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,

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

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

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

In which the term $\mathbf{u}\left(k\right)$ 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. (<-------- week 6)