Advanced Autonomous Systems

Bayesian Estimation- Part 2

In the previous lecture, we concluded that one remarkable aspect of the Bayesian approach is its ability to perform the fusion of different sources of information. Suppose we have a sequence of observations,

$$\begin{aligned} \mathbf{y}_0 &= h_0(\mathbf{x}) + \boldsymbol{\eta}_0 \\ \mathbf{y}_1 &= h_1(\mathbf{x}) + \boldsymbol{\eta}_1 \\ &\cdots \\ \mathbf{y}_M &= h_M(\mathbf{x}) + \boldsymbol{\eta}_M \end{aligned} \qquad \text{(E1)}$$

In which each observation is associated to an observation model $\mathbf{y} = h_i(\mathbf{x})$. The functions $\{h_i(\mathbf{x})\}_{i=0}^{M}$ can be non-linear and diverse, i.e. they may be different, i.e. $h_i(\,) \neq h_j(\,)$). The observations $\{\mathbf{y}_i\}_{i=0}^{M}$ are polluted by noise; those noise components $\{\boldsymbol{\eta}_i\}$ can even have different probability distributions (e.g. when we use multiple and diverse sensors or those are used under different conditions of operation.)
However, there is an assumption about the noise components: we assume they are independent, i.e. $p_{(\boldsymbol{\eta}_i,\boldsymbol{\eta}_j)}(\boldsymbol{\eta}_i,\boldsymbol{\eta}_j) = p_{\boldsymbol{\eta}_i}(\boldsymbol{\eta}_i)\cdot p_{\boldsymbol{\eta}_j}(\boldsymbol{\eta}_j), \quad \forall(i,j) \quad / \quad i \neq j$.

By successively applying the Bayes' rule, we obtain that

$$p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x} \mid \mathbf{Y} = \{\mathbf{y}_i\}_{i=0}^{K}\right) \quad \propto \quad p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y} = \mathbf{y}_K \mid \mathbf{x}\right) \cdot p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x} \mid \mathbf{Y} = \{\mathbf{y}_i\}_{i=0}^{K-1}\right) \qquad \text{(E2)}$$

from which the resulting PDF $p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x} \mid \mathbf{Y} = \{\mathbf{y}_i\}_{i=0}^{M}\right)$ is obtained.

Note that in our notation we use the ***Y*** for denoting a set of measurements, and ***y*** for denoting an individual measurement.

This probability function expresses the conditional probability of the state ***x*** given the set of obtained observations $\{\,\mathbf{y}_i\,\}_{i=0}^{M}$. I.e., our belief about the value of ***x*** based on information provided by the set of measurements of certain system's outputs.
Note that at each instance of (E2) we considered the likelihood function which is associated to an individual measurement, and the prior PDF which was available before processing that particular measurement. At iteration *i* the prior (prior belief about ***x***) is the resulting posterior PDF of iteration *i-1*. Its associated likelihood function is defined by $h_i(.)$ and $p_{\eta_i}(.)$, i.e. that particular observation's function, and the statistics of the uncertainty which is assumed to be polluting that measurement.
This process is usually called *Sensor Data Fusion*, which means: the process of estimating the hidden state ***x*** by processing measurements from different sensors and models. Using different sensors means that we are dealing with different observation models, and with noises of different statistical properties.

Process Model

In Equations (E1) and (E2) we assumed that the "true" state, ***x***, is time invariant, i.e. it does not change during the sequence of observations. However, the system's state is usually time varying; a typical case is when it follows certain dynamics, for instance some differential equation, or a discrete time model such as:

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) \qquad \text{(E3)}$$

In general, our belief about $\mathbf{x}(k+1)$ and $\mathbf{x}(k)$ will be represented by different PDFs, because these variables are different entities. If we perfectly knew the process model, i.e. *F(.)*, and the inputs $\mathbf{u}(k)$, and the PDF about the variable $\mathbf{x}(k)$, then we would be able to obtain the PDF about $\mathbf{x}(k+1)$.
The mathematical description of that operation takes the following expression:

$$p_{\mathbf{x}(k+1)}\left(\mathbf{x}(k+1)\right) = \left(p_{\mathbf{x}(k)}\left(\mathbf{x}\right) \cdot J\left(\mathbf{x},\mathbf{u}\right)^{-1} \right)\Big|_{\mathbf{x}=F^{-1}\left(\mathbf{x}(k+1),\mathbf{u}(k)\right)}$$
$$J\left(\mathbf{x},\mathbf{u}\right) = \left|\frac{\partial F}{\partial \mathbf{x}}\left(\mathbf{x},\mathbf{u}\right)\right| \qquad \text{(E4)}$$

In which we assume the inverse function, of *F(x , u)*, as a function of x, does exist,
$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) \Rightarrow \mathbf{x}(k) = F^{-1}\left(\mathbf{x}(k+1), \mathbf{u}(k)\right)$

The mathematical demonstration of E4 is presented in appendix 2.

Usually, we do not literally apply such a procedure (E4), because there are methods for performing this definition in an efficient way, as we will see later.

As our knowledge about the system's dynamics is usually not perfect, we assume the existence of additive uncertainty,

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \boldsymbol{\xi}(k) \qquad \text{(E5)}$$

This means that this process can be seen as the result of applying (E3), and subsequently adding certain noise to its result, for obtaining (E5). The process of adding uncertainty is described in the following equation,

$$\mathbf{z} = \mathbf{x} + \boldsymbol{\xi} \qquad \text{(E6)}$$

In (E6) we consider that the couple of variables $\mathbf{x}, \boldsymbol{\xi}$ are random variables whose PDFs, $p_{\mathbf{x}}(\mathbf{x})$ and $p_{\xi}(\boldsymbol{\xi})$, are known, and we also consider that both RV's are independent. According to (E6) the variable $\mathbf{z}$ is also a RV, because it is the result of adding two RV's. The resulting RV, $\mathbf{z}$ (which is fully dependent on $\mathbf{x}$ and $\boldsymbol{\xi}$) will be described by a PDF $p_{\mathbf{z}}(\mathbf{z})$. This PDF is well defined by the functions $p_{\mathbf{x}}(.)$ and $p_{\xi}(.)$; it is the result of the following **convolution** operation,

$$p_{\mathbf{z}}(\mathbf{z}) = \int_{\substack{\text{domain} \\ \text{of } \xi}} p_{\mathbf{x}}(\mathbf{z}-\xi)\cdot p_{\xi}(\xi)\cdot d\xi \qquad \text{(E7)}$$

(Note: appendix 2 provides the demonstration of (E7))

Now, we focus our attention to our case, (E5). The process model (expressed in (E5)) transforms the variable $\mathbf{x}(k)$ for obtaining $\mathbf{x}(k+1)$. In the same way, our belief about the variable $\mathbf{x}$ is also transformed, i.e. the original PDF that describes the RV $\mathbf{x}(k)$ is transformed in a new PDF about $\mathbf{x}(k+1)$. Note that $\mathbf{x}(k+1)$ is a new RV; which, in this case, is function of the RVs $\mathbf{x}(k)$ and $\xi(k)$; this is why the PDF of $\mathbf{x}(k+1)$ is a function of the PDFs of $\mathbf{x}(k)$ and $\xi(k)$, and of the process model F*(.)*. In theory, by application of the operations defined in (E4) and (E7), we can obtain the new PDF $p_{\mathbf{x}(k+1)}(\mathbf{x}(k+1))$.

It is clear that the operation in (E6) has the effect of decreasing the quality of the estimates (the resulting PDF is spread, more than the PDF of the original RVs). The Convolution operation will usually diffuse the original PDF $p_{\mathbf{x}}(.)$. In this case, the operation is applied in (E6) ($p_{\mathbf{x}}(.)$ being convolved by $p_{\xi}(.)$ for producing a new PDF $p_{\mathbf{z}}(z)$) for modeling the effect of the additive noise ξ on the original RV $\mathbf{x}$. Our knowledge about the new variable would have lower quality (lower certainty), i.e. the PDF $p_{\mathbf{z}}(.)$ is less *informative* than $p_{\mathbf{x}}(.)$. The exception for this would be if $p_{\xi}(.)$ is represented by a Dirac Delta Function (which means there is actually no added noise).

Bayesian Estimation for a Dynamic System

Consider a system, which follows certain dynamics,

$$\mathbf{x}(k+1) = F(\mathbf{x}(k),\mathbf{u}(k)) + \xi(k)$$

and that some of its outputs are sampled at certain times

$$\mathbf{y}(k) = h(\mathbf{x}(k),k) + \boldsymbol{\eta}(k)$$

In order to fully apply the Bayesian approach, two different operations need to be performed. At time step *k* the process model is applied in order to obtain the PDF about the RV $\mathbf{x}(k)$, based on the PDF about $\mathbf{x}(k-1)$. This operation is called ***prediction***. Then, if some observations are available at time *k*, their associated likelihood functions can be evaluated and applied through the Bayes' rule; consequently, an improved version of the belief about $\mathbf{x}(k)$ is obtained. This step is called ***update***.

At time $(k+1)$ the process model is applied again, for predicting the evolution of the state, i.e. a PDF (about $\mathbf{x}(k+1)$), which is based on the previously updated belief about $\mathbf{x}(k)$. This sequence of predictions and Bayes' rule updates can be applied continuously, to keep producing a belief about $\mathbf{x}$ at the current time.

$$p_{\mathbf{x}(k)|\mathbf{Y}(k-1)}(\mathbf{x}(k)\,|\,\mathbf{Y}(k-1)) = \text{Prediction}\left(p_{\mathbf{x}(k-1)|\mathbf{Y}(k-1)}(\mathbf{x}(k-1)\,|\,\mathbf{Y}(k-1))\right)$$

$$p_{\mathbf{x}(k)|\mathbf{Y}(k)}(\mathbf{x}(k)\,|\,\mathbf{Y}(k)) \;\propto\; p_{\mathbf{y}(k)|\mathbf{x}(k)}(\mathbf{y}(k)\,|\,\mathbf{x}(k))\Big|_{\mathbf{y}(k)=\mathbf{y}_K} \cdot p_{\mathbf{x}(k)|\mathbf{Y}(k-1)}(\mathbf{x}(k)\,|\,\mathbf{Y}(k-1)) \qquad \text{(E8)}$$

The meaning of the notation used in (E8) is the following:

$\mathbf{Y}(i)$: all the observations, which have occurred till time ***i*** **(**including the observations that time as well)

$\mathbf{y}(i)$: the observation (or the set of observations) that occurred at time ***i***,

$\mathbf{y}_i$: measured value of $\mathbf{y}(i)$.

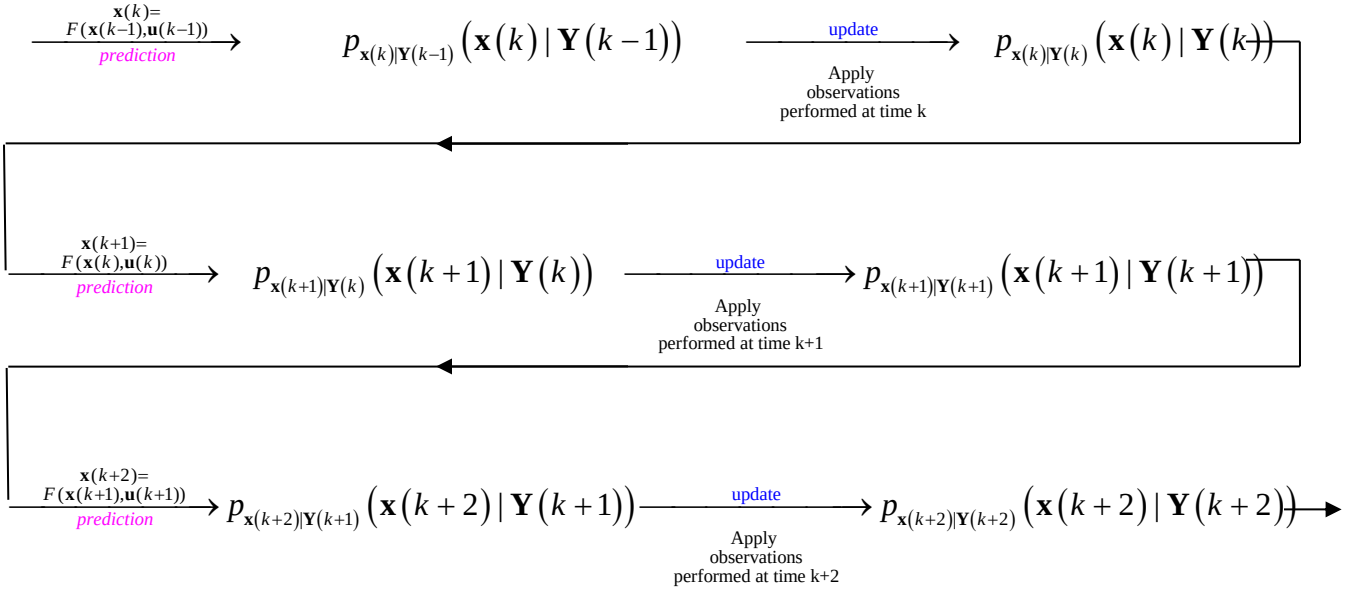
$p_{\mathbf{x}(k-1)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k-1)\,|\,\mathbf{Y}(k-1)\right)$: PDF describing the RV $\mathbf{x}(k-1)$, based on observations collected till time $k-1$.

$p_{\mathbf{x}(k)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k)\,|\,\mathbf{Y}(k-1)\right)$: the PDF describing the RV $\mathbf{x}(k)$, based on observations collected till time $k-1$.

$p_{\mathbf{x}(k)|\mathbf{Y}(k)}\left(\mathbf{x}(k)\,|\,\mathbf{Y}(k)\right)$: the PDF for describing the RV $\mathbf{x}(k)$ based on observations collected till time k.

$p_{\mathbf{y}(k)|\mathbf{x}(k)}\left(\mathbf{y}(k)\,|\,\mathbf{x}(k)\right)\Big|_{\mathbf{y}(k)=\mathbf{y}_K}$: likelihood function based on the measurement $\mathbf{y}_k$, and the observation model of $\mathbf{y}(k)$.

The iterative process can be represented by the following sequence:



(The sequence continues, for subsequent discrete times, $(k+i), \quad i>2$)

This approach is flexible in the sense that the Bayesian update is synthetized based on the likelihood functions associated to the available measurements or observations; however, if at certain step no observation is available then no update is applied at that step, and the subsequent prediction step is applied as usual.
In the next section we will discuss about how to solve, in practice, what we defined mathematically in the previous section. Our mathematical explanation was correct; however, the difficulty is about how to make this approach tractable in a computer, and usually in real-time.

(Note: we will run some simulation, in class, to see this class of estimation process in action)

Bayesian Approach Applied to Gaussian PDFs

In this section we analyze why the Bayesian Estimation process can be simplified, in certain cases, if the involved PDF functions are restricted to be of the Gaussian family. In this section of the lecture notes, we discuss about certain properties and approaches which will be exploited for simplifying the application of the concepts previously discussed; i.e. to make the theory tractable in practical applications.

Linear Operation Applied on Gaussian RVs

Certain usual operations (linear and affine transformations), are easily treated when applied to Gaussian RVs.

1) If we add two independent Gaussian RVs, we obtain a new RV, which is Gaussian, and whose expected value and covariance matrix can be easily evaluated,

$$\begin{gathered} \mathbf{z} = \mathbf{x} + \mathbf{y}, \\ \mathbf{x} \in \mathrm{R}^n, \mathbf{y} \in \mathrm{R}^n, \quad \mathbf{x} \sim N\left(\mathbf{x}\,;\hat{\mathbf{x}},\mathbf{P}_{\mathbf{x}}\right), \quad \mathbf{y} \sim N\left(\mathbf{y}\,;\hat{\mathbf{y}},\mathbf{P}_{\mathbf{y}}\right), \quad \mathrm{cov}\left(\mathbf{x},\mathbf{y}\right) = 0 \\ \Downarrow \\ \mathbf{z} \sim N\left(\mathbf{z}\,;\hat{\mathbf{z}},\mathbf{P}_{\mathbf{z}}\right); \quad \hat{\mathbf{z}} = \hat{\mathbf{x}} + \hat{\mathbf{y}}, \quad \mathbf{P}_{\mathbf{z}} = \mathbf{P}_{\mathbf{x}} + \mathbf{P}_{\mathbf{y}} \end{gathered} \tag{E9}$$

(convolving two Gaussian PDFs will produce a Gaussian PDF whose expected value and covariance matrix are easily evaluated, as expressed in E9)

2) If we apply a linear transformation on a multivariate Gaussian RV, then the resulting variable is a Gaussian RV, whose expected value and covariance matrix can be easily evaluated,

$$\begin{gathered} \mathbf{z} = \mathbf{A} \cdot \mathbf{x} \\ \mathbf{z} = \begin{bmatrix} z_1 & z_2 & .. & z_m \end{bmatrix}^T \in R^m, \quad \mathbf{x} = \begin{bmatrix} x_1 & x_2 & .. & x_n \end{bmatrix}^T \in R^n, \quad \mathbf{A} \in R^{m \times n} \\ \mathbf{x} \sim N\left(\mathbf{x}\,;\hat{\mathbf{x}},\mathbf{P}_{\mathbf{x}}\right) \\ \Downarrow \\ \mathbf{z} \sim N\left(\mathbf{z}\,;\hat{\mathbf{z}},\mathbf{P}_{\mathbf{z}}\right) \\ \hat{\mathbf{z}} = \mathbf{A} \cdot \hat{\mathbf{x}} \\ \mathbf{P}_{\mathbf{z}} = \mathbf{A} \cdot \mathbf{P}_{\mathbf{x}} \cdot \mathbf{A}^T \end{gathered} \tag{E10}$$

(in which **A** is a deterministic matrix, i.e. it is not a function of **X** or another RV.)

Here we use the notation $\mathbf{P}_{\mathbf{x}}$ for indicating the covariance matrix of the RV $\mathbf{x}$ (whose PDF, which we usually denote as $p_{\mathbf{x}}(\mathbf{x})$, is Gaussian). Do not confuse, in our notation, $\mathbf{P}_{\mathbf{x}}$ and $p_{\mathbf{x}}(\mathbf{x})$.

Note that (E10) involves multiplying and adding RV's. The operations for obtaining the resulting parameters $\hat{\mathbf{z}}, \mathbf{P}_{\mathbf{z}}$ has implicitly "applied" all the scaling and convolutions needed, defined in (E4) and (E7). All this simplicity is due to the fact that **x** is a Gaussian RV, and that all the operations are linear. Just imagine doing the convolutions of these n-dimensional PDFs instead of these efficient matrix operations!

First Order Approximation for Non-Linear cases

Consider the more complicated problem, in which **z** is the result of a NON-LINEAR function of a RV **x**.

$$\mathbf{z} = a(\mathbf{x})$$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_m \end{bmatrix} \in R^m, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \in R^n, \quad a(\ ) : R^n \to R^m \qquad \text{(E11)}$$

Usually, a non-linear transformation of the Gaussian RV **X** produces a new RV **Z**, which is not Gaussian. However, under certain conditions, the resulting RV may have a PDF which can be well approximated by a Gaussian PDF (as we will see soon).

Linearization:

A well-known approach, for locally approximating a function $a(\mathbf{x})$ by a linear function, is the first order Taylor expansion of $a(\mathbf{x})$, at a given point $\mathbf{x} = \mathbf{x}_0$.

$$\mathbf{z} = a(\mathbf{x}) \cong a(\mathbf{x}_0) + \mathbf{A} \cdot (\mathbf{x} - \mathbf{x}_0)$$

$$\mathbf{A} = \left. \frac{\partial a(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_0}$$

$$\frac{\partial a(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial a_1}{\partial x_1} & \frac{\partial a_1}{\partial x_2} & \dots & \frac{\partial a_1}{\partial x_n} \\ \frac{\partial a_2}{\partial x_2} & \dots & \dots & \frac{\partial a_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial a_m}{\partial x_1} & \dots & \dots & \frac{\partial a_m}{\partial x_n} \end{bmatrix} \qquad \text{(E12)}$$

In equation (E4), the operation for obtaining the PDF about **z**, provided that ***F(.)*** and the PDF of **x** are known, is well defined and expressed mathematically; however, that operation is expensive when applied. The complexity and cost of applying (E4) can be dramatically reduced by the process which is discussed in the following section. Suppose that **x** is a Gaussian RV and that the function $a(\mathbf{x})$ is "almost linear" (*) in a region of the domain of **x**, in which the PDF of **x** is dominant, i.e. where the PDF $p_{\mathbf{x}}(\mathbf{x})$ concentrates almost all its probability. Then a first order approximation of (E11) would be adequate and useful for obtaining the resulting PDF, $p_{\mathbf{z}}(\mathbf{z})$, expressed as a Gaussian PDF. In order to do it, the procedure is the following one:

(*) It can be satisfactorily approximated in that region, by using a linear model.

$$\mathbf{z} = a(\mathbf{x})$$
$$\mathbf{x} \sim N(\mathbf{x};\ \hat{\mathbf{x}}, \mathbf{P}_\mathbf{x})$$
$$\Downarrow$$
$$\mathbf{z} \sim N(\mathbf{z};\ \hat{\mathbf{z}}, \mathbf{P}_\mathbf{z})$$

(E13)

$$\hat{\mathbf{z}} = a(\hat{\mathbf{x}})$$
$$\mathbf{P}_\mathbf{z} = \mathbf{A} \cdot \mathbf{P}_\mathbf{x} \cdot \mathbf{A}^T$$
$$\mathbf{A} = \left.\frac{\partial a(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}}$$

(note that the value of **x,** at which the derivatives of the linearization are evaluated, is the expected value $\hat{\mathbf{x}}$)

These operations are straightforward, e.g. just the evaluation of partial derivatives, and some matrix operations. The matrix **A** is called the JACOBIAN matrix of the function $a(.)$; it is a matrix, which expresses the partial derivatives of the components of $a(.)$ respect to all the components of the variable **x**.

Example

Suppose this case, in which certain variable **z** is the result of evaluating a function h**(x)**, being **x** is a 3D Gaussian RV. The function is,

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} h_1(x_1, x_2, x_3) \\ h_2(x_1, x_2, x_3) \end{bmatrix} = \begin{bmatrix} \sin\left(1 + \frac{x_1 + 2 \cdot x_2}{100}\right) + 5 \cdot x_3 \\ \left(1 + \frac{x_1}{100}\right)^3 + x_2 \end{bmatrix}$$

This function is applied to the RV, whose estimate is expressed by a Gaussian PDF, which has the following expected value and covariance matrix,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \sim N(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P}_\mathbf{x}), \quad \hat{\mathbf{x}} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{P}_\mathbf{x} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

We know that the PDF of **z** can be obtained (its approximation). It is a Gaussian PDF,

$$\mathbf{z} \sim N(\mathbf{z}; \hat{\mathbf{z}}, \mathbf{P}_\mathbf{z})$$

$$\hat{\mathbf{z}} = h(\hat{\mathbf{x}})$$
$$\mathbf{P}_\mathbf{z} = \mathbf{H} \cdot \mathbf{P}_\mathbf{x} \cdot \mathbf{H}^T$$
$$\mathbf{H} = \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}}$$

Firstly, we evaluate the Jacobian matrix of the function $h(\mathbf{x})$ at the "point of operation" (the expected value $\hat{\mathbf{x}}$).

$$\mathbf{H}=\begin{bmatrix}\frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \frac{\partial h_1}{\partial x_3}\\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \frac{\partial h_2}{\partial x_3}\end{bmatrix}=\begin{bmatrix}\frac{1}{100}\cdot\cos\left(1+\frac{x_1+2\cdot x_2}{100}\right) & \frac{2}{100}\cdot\cos\left(1+\frac{x_1+2\cdot x_2}{100}\right) & 5\\ \frac{3}{100}\cdot\left(1+0.01\cdot x_1\right)^2 & 1 & 0\end{bmatrix}=$$

$$=\begin{bmatrix}\frac{1}{100}\cdot\cos\left(1+\frac{0+2\cdot 2}{100}\right) & \frac{2}{100}\cdot\cos\left(1+\frac{0+2\cdot 2}{100}\right) & 5\\ \frac{3}{100}\cdot\left(1+0.01\cdot 0\right)^2 & 1 & 0\end{bmatrix}=$$

$$\mathbf{H}=\begin{bmatrix}\frac{1}{100}\cdot\cos(1.04) & \frac{2}{100}\cdot\cos(1.04) & 5\\ \frac{3}{100} & 1 & 0\end{bmatrix}=\begin{bmatrix}0.0051 & 0.0102 & 5\\ 0.03 & 1 & 0\end{bmatrix}$$

Now, we use **H** for evaluating the covariance of **Z**. The expected value of ***Z*** and the covariance matrix are then:

$$\hat{\mathbf{z}}=\begin{bmatrix}\sin(1.04)+5\cdot 1\\ \left(1+\frac{0}{100}\right)^3+2\end{bmatrix}=\begin{bmatrix}5.9624\\ 3\end{bmatrix}$$

$$\mathbf{P}_{\mathbf{z}}=\begin{bmatrix}0.0051 & 0.0102 & 5\\ 0.03 & 1 & 0\end{bmatrix}\cdot\begin{bmatrix}2 & 1 & 1\\ 1 & 4 & 1\\ 1 & 1 & 2\end{bmatrix}\cdot\begin{bmatrix}0.0051 & 0.0102 & 5\\ 0.03 & 1 & 0\end{bmatrix}^T=$$

$$=\begin{bmatrix}0.0051 & 0.0102 & 5\\ 0.03 & 1 & 0\end{bmatrix}\cdot\begin{bmatrix}2 & 1 & 1\\ 1 & 4 & 1\\ 1 & 1 & 2\end{bmatrix}\cdot\begin{bmatrix}0.0051 & 0.03\\ 0.0102 & 1\\ 5 & 0\end{bmatrix}=\begin{bmatrix}50.15 & 6.55\\ 6.55 & .4.78\end{bmatrix}_{(2\times 2\text{ matrix})}$$

This matrix multiplication will produce a matrix of size *2 x 2*, i.e. the covariance matrix of the RV **Z**. Note that the sizes of the matrixes are consistent. $\mathbf{P}_{\mathbf{X}}$ is a *3x3* matrix (because it is the covariance of a 3D RV). $\mathbf{P}_{\mathbf{z}}$ is a *2x2* matrix (**Z** is a 2D RV). The Jacobian matrix **H** is *2 x 3* because it is the Jacobian matrix of a function that transforms (in this example) a 3D variable to a 2D one (i.e. $h(.):R^3\rightarrow R^2$).

Why are we discussing about these matters? We are discussing about certain mathematical facts. Based on the first order Taylor approximations and on the Bayes' rule, we have some of the fundamentals for justifying the *Extended Kalman Filter* (**EKF**), which will be discussed later in our lectures.

Now we focus, again, on the Process model expressed in (E3),

$$\mathbf{x}(k+1)=F\left(\mathbf{x}(k),\mathbf{u}(k)\right)$$

If $F(.)$ and $\mathbf{u}(k)$ are well known and $\mathbf{x}(k)$ is a Gaussian RV, which has well known expected value and covariance matrix, then we are able to obtain an approximation for the PDF of $\mathbf{x}(k+1)$. This approximate PDF

is Gaussian, and its parameters are evaluated through application of (E13). (where $\mathbf{x}(k+1)$, $\mathbf{x}(k)$, $F(.)$ take the roles of the previously used $\mathbf{z}, \mathbf{x}, a(.)$, respectively.)

Additive noise in the Prediction Step:

Could we deduce how to solve the following case?

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \xi(k),$$

in which the noise $\xi(k)$ is assumed to be white Gaussian noise, with covariance $\mathbf{P}_{\xi(k)}$,

$$\xi(k) \sim N\left(\xi(k); \bar{\mathbf{0}}, \mathbf{P}_{\xi(k)}\right)$$

(Hint: The problem involves the addition of two independent Gaussian RV's!)

We can also consider an approximate solution for the more general case $\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k), \xi(k)\right)$, in which the noise is not simply additive. This could be due to uncertainty present in the inputs $\mathbf{u}(k)$, i.e.: $\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k) + \xi_{\mathbf{u}}(k)\right)$ where $\xi_{\mathbf{u}}(k)$ is uncertainty in our knowledge about the inputs $\mathbf{u}(k)$.

(Hint: the function could be linearized with respect to the argument **u***(k)*).

Finally, we can also analyze the case, in which certain noise does pollute the input of the system.

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k) + \xi_{\mathbf{u}}(k)\right) + \xi(k)$$

In which the three RVs, $\mathbf{X}(k), \xi_{\mathbf{u}}(k), \xi(k)$, are independent Gaussian RVs.

(to be discussed during the lecture).

Appendix 1: Demonstration: PDF of the addition of two independent RVs.

Consider the following constraint: $\mathbf{z} = \mathbf{x} + \mathbf{y}$, in which **x** and **y** are two independent random variables, whose joint PDF is $p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})$. Consequently, the Conditional PDF of the dependent RV **z**, given $(\boldsymbol{x}, \boldsymbol{y})$, is a Dirac Delta function (because the relation $\mathbf{z} = \mathbf{x} + \mathbf{y}$ is exactly defined)

$$p_{\mathbf{z}|\mathbf{x},\mathbf{y}}(\mathbf{z}|\mathbf{x},\mathbf{y}) = \delta(\mathbf{z} - \mathbf{x} - \mathbf{y})$$

which implies that the Joint PDF, for the three random variables**,** $(\mathbf{x},\mathbf{y},\mathbf{z})$**,** is then

$$p_{\mathbf{x},\mathbf{y},\mathbf{z}}(\mathbf{x},\mathbf{y},\mathbf{z}) = p_{\mathbf{z}|\mathbf{x},\mathbf{y}}(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = \delta(\mathbf{z} - \mathbf{x} - \mathbf{y}) \cdot p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})$$

Now we can evaluate $p_{\mathbf{z}}(\mathbf{z})$, by marginalizing $p_{\mathbf{x},\mathbf{y},\mathbf{z}}(\mathbf{x},\mathbf{y},\mathbf{z})$ with respect to the RVs $(\mathbf{x},\mathbf{y})$,

$$p_{\mathbf{z}}(\mathbf{z}) = \int_{\Omega_{\mathbf{x},\mathbf{y}}} p_{\mathbf{x},\mathbf{y},\mathbf{z}}(\mathbf{x},\mathbf{y},\mathbf{z}) \cdot d\mathbf{x} \cdot d\mathbf{y} = \int_{\Omega_{\mathbf{x},\mathbf{y}}} p_{\mathbf{z}|\mathbf{x},\mathbf{y}}(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) \cdot d\mathbf{x} \cdot d\mathbf{y}$$

then

$$p_{\boldsymbol{z}}(\boldsymbol{z}) = \int_{\Omega_{\mathbf{x},\mathbf{y}}} \delta(\mathbf{z} - \mathbf{x} - \mathbf{y}) \cdot p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) \cdot d\mathbf{x} \cdot d\mathbf{y} =$$

$$= \int_{\Omega_{\mathbf{y}}} \int_{\Omega_{\mathbf{x}}} \delta(\mathbf{z} - \mathbf{x} - \mathbf{y}) \cdot p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) \cdot d\mathbf{x} \cdot d\mathbf{y}$$

(Note: we use the notation $\Omega_{\mathbf{W}}$ to indicate the full domain of the variable **W**.)

Because the random variables **x** and **y** are assumed to be independent, we can factorize their joint PDF $p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})$, $\Rightarrow p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = p_{\mathbf{x}}(\mathbf{x}) \cdot p_{\mathbf{y}}(\mathbf{y})$. Consequently,

$$p_{\mathbf{z}}(\mathbf{z}) = \int_{\Omega_{\mathbf{y}}} \left(\int_{\Omega_{\mathbf{x}}} \delta(\mathbf{z} - \mathbf{x} - \mathbf{y}) \cdot p_{\mathbf{x}}(\mathbf{x}) \cdot d\mathbf{x} \right) \cdot p_{\mathbf{y}}(\mathbf{y}) \cdot d\mathbf{y}$$

Now we evaluate the inner integral

$$\int_{\Omega_{\mathbf{x}}} \delta(\mathbf{z} - \mathbf{x} - \mathbf{y}) \cdot p_{\boldsymbol{x}}(\mathbf{x}) \cdot d\mathbf{x} = p_{\mathbf{x}}(\mathbf{z} - \mathbf{y}),$$

which implies that

$$p_{\mathbf{z}}(\mathbf{z}) = \int_{\Omega_{\mathbf{y}}} p_{\mathbf{x}}(\mathbf{z} - \mathbf{y}) \cdot p_{\mathbf{y}}(\mathbf{y}) \cdot d\mathbf{y}$$

This integral operation is a convolution operation, involving the PDFs of both independent RVs.

(You may commute the order of integration, to be $d\mathbf{y} \cdot d\mathbf{x}$ in place of $d\mathbf{x} \cdot d\mathbf{y}$, and you would get an equivalent convolution operation, that would produce the same result, i.e. $p_{\mathbf{z}}(\mathbf{z})$)

Note: This is just FYI. There is no need to memorize this demonstration (although if you understand it, you can reproduce it at any time).

Appendix 2 : PDF of a transformed RV.

If given an RV $\mathbf{a}$, whose PDF is $p_{\mathbf{a}}(\mathbf{a})$, we obtain a transformed version, RV $\mathbf{b} = F(\mathbf{a})$, in which $\mathbf{a} \in \mathbf{R}^n, \mathbf{b} \in \mathbf{R}^n$; $F: \mathbf{R}^n \rightarrow \mathbf{R}^n$, and in which the function F(.) has inverse, and whose Jacobian matrix $\frac{\partial F(\mathbf{a})}{\partial \mathbf{a}}$ does exist.
We want to obtain the PDF which describes the statistical properties of the RV $\mathbf{b}$.

As the RV $\mathbf{b}$ is exactly defined by the RV $\mathbf{a}$, then we can define its conditional PDF as follows,

$$p_{\mathbf{b}|\mathbf{a}}(\mathbf{b}|\mathbf{a}) = \delta\big(\mathbf{b} - F(\mathbf{a})\big)$$

in which $\delta(\ \ )$ is the Dirac's delta function. This is because $\mathbf{b}$ is perfectly (i.e. deterministically) defined by $\mathbf{a}$ (so that if we know the value of $\mathbf{a}$, we perfectly know the value of $\mathbf{b}$.)
Consequently, the joint PDF about both RVs can be expressed as follows,

$$p_{\mathbf{a},\mathbf{b}}(\mathbf{a},\mathbf{b}) = p_{\mathbf{b}|\mathbf{a}}(\mathbf{b}|\mathbf{a}) \cdot p_{\mathbf{a}}(\mathbf{a})$$

resulting in

$$p_{\mathbf{a},\mathbf{b}}(\mathbf{a},\mathbf{b}) = \delta\big(\mathbf{b} - F(\mathbf{a})\big) \cdot p_{\mathbf{a}}(\mathbf{a})$$

In order to obtain the marginal PDF $p_{\mathbf{b}}(\mathbf{b})$, we proceed marginalizing the joint PDF $p_{\mathbf{a},\mathbf{b}}(\mathbf{a},\mathbf{b})$, as it follows,

$P_{\mathbf{b}}(\mathbf{b}) = \int_{\Omega_{\mathbf{a}}} p_{\mathbf{a},\mathbf{b}}(\mathbf{a},\mathbf{b}) \cdot d\mathbf{a} = \int_{\Omega_{\mathbf{a}}} \cdot p_{\mathbf{b}|\mathbf{a}}(\mathbf{b}|\mathbf{a}) \cdot p_{\mathbf{a}}(\mathbf{a}) \cdot d\mathbf{a} = \int_{\Omega_{\mathbf{a}}} \cdot \delta\big(\mathbf{b} - F(\mathbf{a})\big) \cdot p_{\mathbf{a}}(\mathbf{a}) \cdot d\mathbf{a}$
(in which $\Omega_{\mathbf{a}}$ means the full domain of the variable $\mathbf{a}$)

For easily integrating it, we apply the change of variables $\boldsymbol{\mu} = F(\mathbf{a})$,

$$P_{\mathbf{b}}(\mathbf{b}) = \int_{\Omega_{\boldsymbol{\mu}}} \cdot \delta(\mathbf{b} - \boldsymbol{\mu}) \cdot p_{\mathbf{a}}(\ F^{-1}(\boldsymbol{\mu})\ ) \cdot J\big(\ F^{-1}(\boldsymbol{\mu})\big) \cdot d\boldsymbol{\mu}$$

in which the scaling factor $J(\mathbf{a})$ is $\left|\frac{\partial F(\mathbf{a})}{\partial \mathbf{a}}\right|$, the determinant of the Jacobian matrix of the transformation (as you know from multi variable differential calculus, in which you applied transformations for simplifying certain double or triple integrals.).

Due to the factor $\delta(\mathbf{b} - \boldsymbol{\mu})$, the integration simply results in:

$$P_{\mathbf{b}}(\mathbf{b}) = p_{\mathbf{a}}(\ F^{-1}(\mathbf{b})\ ) \cdot J\big(\ F^{-1}(\mathbf{b})\big)$$

We can see that the PDF about $\mathbf{b}$, $P_{\mathbf{b}}(\mathbf{b})$, is clearly defined by the PDF $P_{\mathbf{a}}(\mathbf{a})$ and by the function $F(\mathbf{a})$.
Now, we simply consider the case in which RV $\mathbf{a}$ is $\mathbf{x}(k)$, and the RV $\mathbf{b}$ is $\mathbf{x}(k+1)$; which corresponds to a prediction step based on our process model; to finally obtain equation E4.