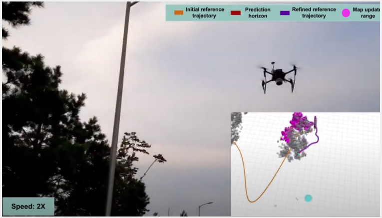
# Autonomous Mobile Robotics

## Kalman Filter

Geesara Kulathunga

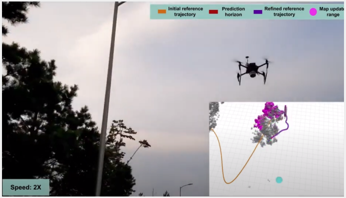



February 28, 2023

# Kalman Filter

# Contents

- Gaussian Distribution
- One Dimensional Kalman Filter
- Multivariate Density Function
- Marginal Density Function
- Multivariate Normal Function
- Two Dimensional Gaussian
- Multiple Random Variable
- Multidimensional Kalman Filter
- Sensor Fusion
- Linearization, Taylor Series Expansion, Linear Systems
- Extended Kalman Filter (EKF)
- Comparison between KF and EKF



## Example 00

Gaussian comes under continuous probability distribution which describes with two parameters: the mean $\mu$ and the variance $\sigma^2$.

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{1}$$

- Let's plot the probability density function (pdf) of $\mathcal{N}(10, 1) \times \mathcal{N}(10, 1)$. What can you say about this result (Fig. 1)?
- Similarly what about the result pdf with $\mathcal{N}(10.2, 1) \times \mathcal{N}(9.7, 1)$ and $\mathcal{N}(8.5, 1.5) \times \mathcal{N}(10.2, 1.5)$? (Fig. 2 and Fig. 3)

## Example 00

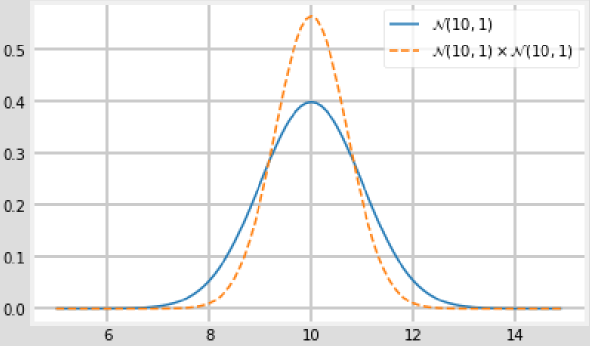



**Figure:** $\mathcal{N}(10, 1) \times \mathcal{N}(10, 1)$

## Example 00

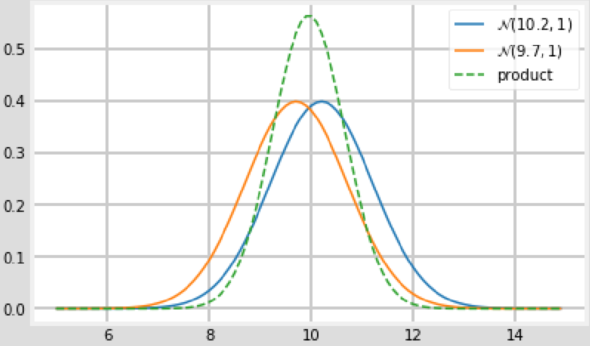



**Figure:** $\mathcal{N}(10.2,\ 1) \times \mathcal{N}(9.7,\ 1)$

## Example 00

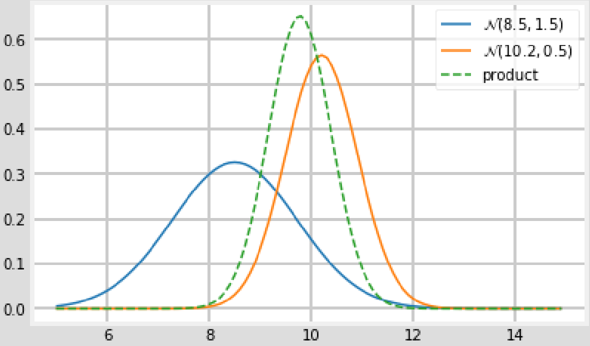



**Figure:** $\mathcal{N}(10.2, 1.5) \times \mathcal{N}(8.5, 1.5)$

### Example 01

There are two independent estimates of the variable $x_1 \sim N(E(x_1), \sigma_1^2)$ and , $x_2 \sim N(E(x_2), \sigma_2^2)$. How can we find the optimal linear combination of these two estimates that represents the corresponding optimal state estimate $\hat{x}$?

## Example 01

- The optimal value of the variable x is assumed to be a linear combination of two estimates

$$\hat{x} = w_1 x_1 + w_2 x_2, \ w_1 + w_2 = 1$$

## Example 01

- The optimal value of the variable x is assumed to be a linear combination of two estimates

$$\hat{x} = w_1 x_1 + w_2 x_2,\; w_1 + w_2 = 1$$

- Let us find out variance of $\hat{x}$

$$\begin{aligned} \sigma^2 &= E\{(\hat{x} - E\{\hat{x}\})^2\} \\ &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2E\{(x_1 - E\{x_1\})(x_2 - E\{x_2\})\} \\ &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 = (1-w)^2\sigma_1^2 + w^2\sigma_2^2 \end{aligned} \tag{2}$$

## Example 01

- The optimal value of the variable x is assumed to be a linear combination of two estimates

$$\hat{x} = w_1 x_1 + w_2 x_2,\; w_1 + w_2 = 1$$

- Let us find out variance of $\hat{x}$

$$\begin{aligned} \sigma^2 &= E\{(\hat{x} - E\{\hat{x}\})^2\} \\ &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2E\{(x_1 - E\{x_1\})(x_2 - E\{x_2\})\} \\ &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 = (1-w)^2\sigma_1^2 + w^2\sigma_2^2 \end{aligned} \tag{2}$$

- Let's try to minimize the variance

$$\frac{\partial\sigma^2}{\partial w} = -2(1-w)\sigma_1^2 + 2w\sigma_2^2 = 0 \tag{3}$$

## Example 01

- The optimal value for w

$$w = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

## Example 01

- The optimal value for w

$$w = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

- The optimal value for $\sigma^2$

$$\sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

## Example 01

- The optimal value for w

$$w = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

- The optimal value for $\sigma^2$

$$\sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

- The optimal value $\hat{x}$

$$\hat{x} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2}$$

## Example 02

Consider example 01, if control input $u(k)$, which is having $\sigma_u^2(k)$ uncertainty, applied to the estimated state $\hat{x}(k)$ determine the uncertainty of the updated state, i.e., $\hat{x}(k+1) = \hat{x}(k) + u(k)$

## Example 02

Consider example 01, if control input $u(k)$, which is having $\sigma_u^2(k)$ uncertainty, applied to the estimated state $\hat{x}(k)$ determine the uncertainty of the updated state, i.e., $\hat{x}(k+1) = \hat{x}(k) + u(k)$

$$\begin{aligned} \sigma^2(k+1) &= E\{(\hat{x}(k+1) - E\{\hat{x}(k+1)\})^2\} \\ &= E\{(\hat{x}(k) - E\{\hat{x}(k)\} + \hat{u}(k) - E\{\hat{u}(k)\})^2\} \\ &= \sigma^2(k) + \sigma_u^2(k) \end{aligned} \tag{4}$$

Let's start off with this example:
`http://david.wf/kalmanfilter/`

When there are more than one random variable, i.e. $\mathbf{x}, \mathbf{y}$, we stack them into a vector, i.e. $[\mathbf{x}, \mathbf{y}]$, and let new random variable be $\mathbf{z} \in \mathbb{R}^n$. Thus, probability density function is called the **Joint Density Function**, which we can express as:

$$p_Z(\mathbf{z}) : \mathbb{R}^n \to \mathbb{R}^+ \tag{5}$$

Subsequently, when **z** takes values with a range, e.g., a and b, the corresponding probability is given by

$$P_z(a \leq \mathbf{z} \leq b) = \int_{a_n}^{b_n} \ldots \int_{a_1}^{b_1} p_z(\mathbf{z}) dz_1..dz_n, \tag{6}$$

that is called the **Marginal Density Function**. So what is the condition should hold for a given join density function to be valid?

$$P_{XY}(x, y) \geq 0, \quad \sum_{\forall x} \sum_{\forall y} P_{XY}(x, y) = 1$$

Moreover, what is the relationship between joint density function and marginal density function?
Consider the following example:

$$P(x, y) = 4x^2 y, \ \ 0 < x < 4, \ \ 0 < y < 3 \tag{7}$$

How can you find the $P_X(x)$?

**Multivariate** means multiple variables. Main intuition here is to represent multiple variables with normal distribution. Let's recap a few basic concepts. The covariance between $x$ and $y$ is

$$cov(\mathbf{x}, \mathbf{y}) = \sigma_{xy} = \mathbb{E}\left[(\mathbf{x} - \mu_x)(\mathbf{y} - \mu_y)^T\right]$$

where $\mathbb{E}[\mathbf{x}]$ is the **expected value** of $\mathbf{x}$ is given by

$$\mathbb{E}[\mathbf{x}] = \begin{cases} \sum_{i=1}^{n} p_i x_i & \text{discrete} \\ \int_{-\infty}^{\infty} f(x)\, x & \text{continuous} \end{cases}$$

If the each data point is equally likely, so the probability of each event is $\frac{1}{N}$. Then the expectation can be express as follows for the discrete case.

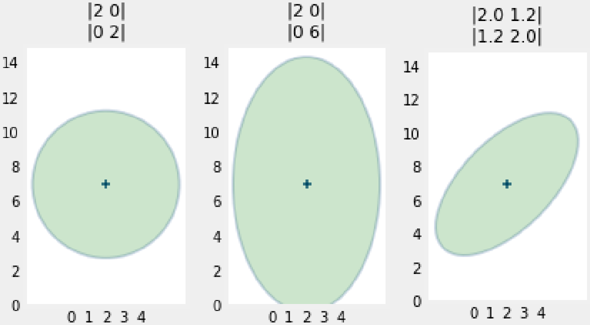
$$\mathbb{E}[\mathbf{x}] = \frac{1}{N} \sum_{i=1}^{n} x_i$$

Compare covariance with variance

$$var(\mathbf{x}) = \sigma_x^2 = \mathbb{E}[(\mathbf{x} - \mu)^2]$$
$$cov(\mathbf{x}, \mathbf{y}) = \sigma_{xy} = \mathbb{E}\big[(\mathbf{x} - \mu_x)(\mathbf{y} - \mu_y)\big] = \mathbb{E}[\mathbf{xy}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]$$

What if $\mathbb{E}[\mathbf{xy}] == \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]$? When there is **no covariance between two random variables**, e.g., $\mathbf{x}, \mathbf{y}$, we can say that those two random variables are uncorrelated, i.e, no linear dependency, but does not say anything about independence. However, if two random variables are independent, then those are uncorrelated, i.e., $E[\mathbf{xy}] == \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]$ as well. Covariance matrix ($\Sigma$) denotes covariances of a multivariate normal distribution.

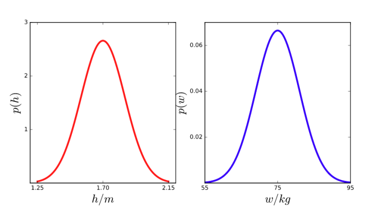
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$




The correlation between two variables can be given a numerical value with **Pearson's Correlation Coefficient** $\rho_{xy}$. It is defined as:

$$\rho_{xy} = \frac{cov(\mathbf{x}, \mathbf{y})}{\sigma_x \sigma_y} \tag{8}$$

Consider the students' weight and height distributions. The marginal density of heights is given by 1.7*m* and variance is 0.025 approximately. Similarly, the marginal density of weights is given by 75kg and variance is 36.

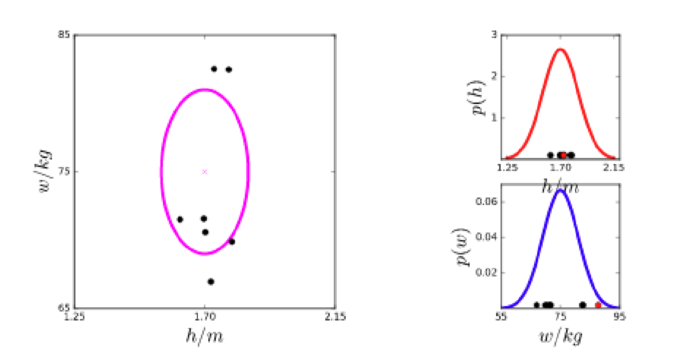



Let's assume there is **no correlation between weight and height distributions**. In other words, student height does not depend on her/his weight. Whats is the **joint probability distribution** p(w, h) ?

Let's assume there is no correlation between weight and height distributions. In other words, student height does not depend on her/his weight.

$$
\begin{aligned}
p(\mathbf{z}) = p(w,h) &= p(w)p(h) \\
p(w,h) &= \frac{1}{\sqrt{2\pi\sigma_1^2}\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{(w-\mu_1)^2}{\sigma_1^2} + \frac{(h-\mu_2)^2}{\sigma_2^2}\right)\right) \\
= \frac{1}{\sqrt{2\pi\sigma_1^2 2\pi\sigma_2^2}} &\exp\left(-\frac{1}{2}\left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)^\top \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)\right) \\
p(\mathbf{z}) &= \frac{1}{\det 2\pi\mathbf{D}^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^\top \mathbf{D}^{-1}(\mathbf{z}-\mu)\right)
\end{aligned} \tag{9}
$$

# Two Dimensional Gaussian

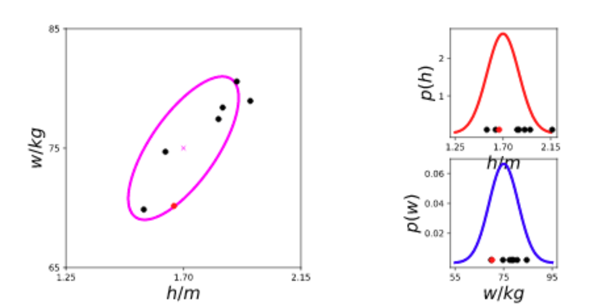


But, are weight and height uncorrelated each other in reality? Answer is no, relation between weight and height is given by $BMI = \frac{w}{h^2}$. How can we incorporate this correlation into the model? Correlation can be added by using the following trick.

$$
\begin{aligned}
p(\mathbf{z}) &= \frac{1}{\det 2\pi \mathbf{D}^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^\top \mathbf{D}^{-1}(\mathbf{z}-\mu)\right) \\
p(\mathbf{z}') &= \frac{1}{\det 2\pi \mathbf{D}^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{R}^\top \mathbf{z}-\mathbf{R}^\top \mu)^\top \mathbf{D}^{-1}(\mathbf{R}^\top \mathbf{z}-\mathbf{R}^\top \mu)\right) \qquad (10) \\
p(\mathbf{z}') &= \frac{1}{\det 2\pi \mathbf{D}^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^\top \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^\top(\mathbf{z}-\mu)\right)
\end{aligned}
$$

where inverse of the covariance matrix is given by $C^{-1} = \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^T$.

$$p(\mathbf{z}) = \frac{1}{\det 2\pi\mathbf{C}^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^{\top}\mathbf{C}^{-1}(\mathbf{z}-\mu)\right) \tag{11}$$



Recall the equation for the normal distribution:

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}(x-\mu)^2/\sigma^2\right] \tag{12}$$

Moreover, multivariate normal function is given by

$$f(\mathbf{z},\ \mu,\ \Sigma) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\ \exp\left[-\frac{1}{2}(\mathbf{z}-\mu)^\mathsf{T}\Sigma^{-1}(\mathbf{z}-\mu)\right] \tag{13}$$

Let $\mathbf{z} = \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}$ be independent random vector of two random variables and their means are given by

$$m_z = \begin{bmatrix} m_x \\ m_y \end{bmatrix} \tag{14}$$

Can you define the covariance matrix whose $\mathbf{z}$?

$$\Sigma = E[(\mathbf{z}-m_z)(\mathbf{z}-m_z)^T] = \begin{bmatrix} \Sigma_{xx} & \Sigma_{x,y} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \tag{15}$$

Now assume **z** is a Gaussian random variable. Thus, how can we define the joint probability density function of **z**?

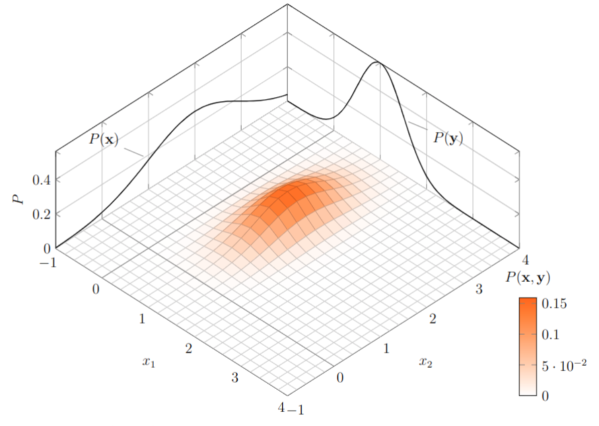
$$p(\mathbf{z}) = p(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{2\pi}^n \sqrt{det\Sigma}} exp(\frac{-1}{2} \begin{bmatrix} \mathbf{x} - m_x \\ \mathbf{y} - m_y \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x} - m_x \\ \mathbf{y} - m_y \end{bmatrix}) \tag{16}$$

Can you try to define $\Sigma_{xy}$ and $\Sigma_{yy}$? Also, what is the different between covariance and variance?

$$\Sigma_{xy} = E[(\mathbf{x} - m_x)(\mathbf{y} - m_y)^T] = E[(\mathbf{y} - m_y)(\mathbf{x} - m_x)^T]^T = \Sigma_{yx}^T \tag{17}$$

Now let's define the conditional mean and variance of $x$ and $y$.

$$\begin{aligned} m_{X|y} &= m_X + \Sigma_{XY}\Sigma_{YY}^{-1}[y - m_Y] \\ \Sigma_{X|y} &= \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{aligned} \tag{18}$$



Let's assume $\Sigma^{-1}$ and $\mu$ are given as

$$\Sigma^{-1} = \begin{bmatrix} 3.5 & 2.5 \\ 2.5 & 4.0 \end{bmatrix}, \quad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{19}$$

To visualize $\Sigma^{-1}$ we have to perform the eigen decomposition of the covariance matrix $\Sigma^{-1}$. Eigen decomposition can be seen as a connection between linear transformation and the covariance matrix. An eigen vector is a vector whose direction remains unchanged when a linear transformation is applied $Av = \lambda v$, where v and $\lambda$ are the the eigenvector and eigenvalue of A. In this example, we have 2 eiegenvectors V and corresponding eigenvalues as digonal matrix L, we can form the $\Sigma^{-1} = C$

$$\begin{aligned} CV &= VL \\ C &= VLV^{-1} \end{aligned} \tag{20}$$

We can use SVD to find out V and L

The optimization criterion used in one dimensional Kalman filter is minimization of least square error of the random variable x. For a given random process, relationship between current and next state and the observation model:

$$\begin{aligned} x_{k+1} &= \Phi_k x_k + w_k \\ z_k &= H_k x_k + v_k \end{aligned} \tag{21}$$

where $x_k$ = (nx1) process state vector at time $t_k$
$\Phi_k$ = (nxn) is the state transition matrix from state $t_k$ to $t_{t+1}$
$w_k$ = (nx1) vector, a white noise sequence with known covariance structure
$z_k$ = (mx1) measurment vector at time $t_k$
$H_k$ = (mxn) transition matrix between measurement and state vector at time $t_k$
$v_k$ = (mx1) measurment error (can you guess) and having zero crosscorrelation with the $w_k$

Corresponding covariance matrices for the $w_k$ and $v_k$ are given by

$$E[w_k w_i^T] = \begin{cases} Q_k & \text{if } i = k \\ 0 & i \neq k \end{cases} \tag{22}$$

$$E[v_k v_i^T] = \begin{cases} R_k & \text{if } i = k \\ 0 & i \neq k \end{cases} \tag{23}$$

$$E[w_k v_i^T] = 0 \text{ for all k and i} \tag{24}$$

- A normal multivariate Gaussian distribution

$$p(\mathbf{x}) = f(\mathbf{x},\ \mu,\ \Sigma) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\ \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right]$$

- Kalman filter equations can be written as

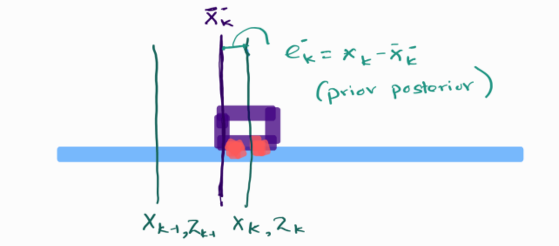
$$p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{u}_k) = \frac{1}{\sqrt{(2\pi)^n|Q|}}$$

$$\exp\left[-\frac{1}{2}(\mathbf{x}_k-(A\mathbf{x}_{k-1}+B\mathbf{u}_{k-1}))^{\mathsf{T}}Q^{-1}(\mathbf{x}_k-(A\mathbf{x}_{k-1}+B\mathbf{u}_{k-1}))\right]$$

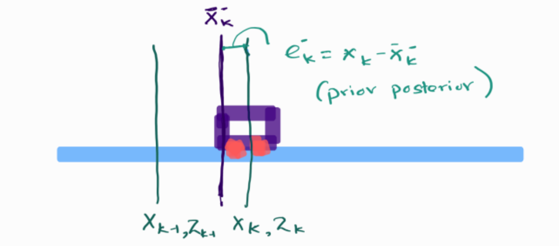
$$p(\mathbf{z}_k|\mathbf{x}_k) = \frac{1}{\sqrt{(2\pi)^n|R|}}\ \exp\left[-\frac{1}{2}(\mathbf{z}_k-\mathbf{x}_k)^{\mathsf{T}}R^{-1}(\mathbf{z}_k-\mathbf{x}_k)\right]$$

where discrete system dynamic is given by
$\mathbf{x}_k = A\mathbf{x}_{k-1} + B\mathbf{u}_{k-1}$

- In order to define the error in the estimation at $t_k$, estimation up to $t_k$ can be defined as $\bar{x}_k^-$.



- In order to define the error in the estimation at $t_k$, estimation up to $t_k$ can be defined as $\bar{x}_k^-$.




- Along with that, estimation error can be defined as:

$$e_k^- = x_k - \bar{x}_k^- \tag{25}$$

- Then the associated error covariance matrix can be defined as

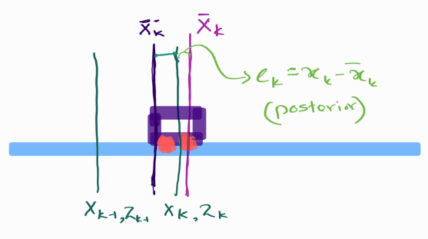
$$P_k^- = E[e_k^- {e_k^-}^T] = E[(x_k - \bar{x}_k^-)(x_k - \bar{x}_k^-)^T] \tag{26}$$

- Then the associated error covariance matrix can be defined as

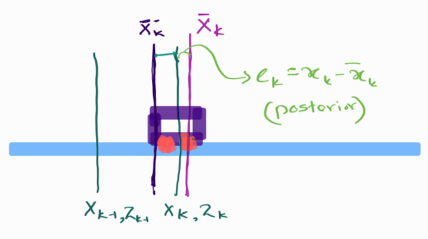
$$P_k^- = E[e_k^- {e_k^-}^T] = E[(x_k - \bar{x}_k^-)(x_k - \bar{x}_k^-)^T] \tag{26}$$

- Let's begin with when there are no prior measurements. What can you say about the process mean and initial estimation?

- Then the associated error covariance matrix can be defined as
$$P_k^- = E[e_k^- e_k^{-T}] = E[(x_k - \bar{x}_k^-)(x_k - \bar{x}_k^-)^T] \tag{26}$$
- Let's begin with when there are no prior measurements. What can you say about the process mean and initial estimation?
- With the assumption of a prior estimation ($\bar{x}_k^-$), in order to improve the prior estimate, it is needed to fuse measurement, $z_k$. Thus, improved estimation can be defined as
$$\bar{x_k} = \bar{x}_k^- + K_k(z_k - H_k\bar{x}_k^-) \tag{27}$$
where $\bar{x_k}$ is the updated (posteriori) estimate and $K_k$ is yet to be determined. When we have the prior information where updated estimation is known, error covariance matrix associated with the updated estimation ($\bar{x_k}$),




- $$P_k = E[e_k {e_k}^T] = E[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T] \tag{28}$$




- $$P_k = E[e_k {e_k}^T] = E[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T] \tag{28}$$
- After considering Eq.21 to Eq. 28 while noting the $(x_k - \bar{x}_k^-)$ which is prior estimation error that is uncorrelated with the current measurement error $v_k$, $P_k$ can be derived as bellow:

$$P_k = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T \tag{29}$$

where I is an identity matrix.

- The error covariance matrix ($P_k$) is minimized for the elements of the state vector being estimated which defined along with the major diagonal of $P_k$. Before minimizing the estimated error, let define two matrix differentiation formulas:

$$\frac{d[trace(AB)]}{dA} = B^T \text{ (A and B must be square)} \tag{30}$$

$$\frac{d[trace(ACA^T)]}{dA} = 2AC \text{ (C must be symmetric)} \tag{31}$$

# Multidimensional Kalman Filter

- The error covariance matrix ($P_k$) is minimized for the elements of the state vector being estimated which defined along with the major diagonal of $P_k$. Before minimizing the estimated error, let define two matrix differentiation formulas:

$$\frac{d[trace(AB)]}{dA} = B^T \text{ (A and B must be square)} \tag{30}$$

$$\frac{d[trace(ACA^T)]}{dA} = 2AC \text{ (C must be symmetric)} \tag{31}$$

- After solving:

$$\frac{d(traceP_k)}{dK_k} = 0 \tag{32}$$

- The error covariance matrix ($P_k$) is minimized for the elements of the state vector being estimated which defined along with the major diagonal of $P_k$. Before minimizing the estimated error, let define two matrix differentiation formulas:

$$\frac{d[trace(AB)]}{dA} = B^T \text{ (A and B must be square)} \tag{30}$$

$$\frac{d[trace(ACA^T)]}{dA} = 2AC \text{ (C must be symmetric)} \tag{31}$$

- After solving:

$$\frac{d(traceP_k)}{dK_k} = 0 \tag{32}$$

- Finally we can find the optimal gain ($K_k$),

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \tag{33}$$

- After substituting the optimal gain($K_k$), covarinace matrix related with is given as

$$P_k = P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} H_k P_k^- \tag{34}$$

or

$$P_k = (I - K_k H_k) P_k^- \tag{35}$$

- After substituting the optimal gain($K_k$), covarinace matrix related with is given as

$$P_k = P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} H_k P_k^- \tag{34}$$

or

$$P_k = (I - K_k H_k) P_k^- \tag{35}$$

- So what can you say about the Kalman filter estimation if the Kalman gain is not an optimal gain?

# Multidimensional Kalman Filter

- After substituting the optimal gain($K_k$), covarinace matrix related with is given as

$$P_k = P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} H_k P_k^- \tag{34}$$

or

$$P_k = (I - K_k H_k) P_k^- \tag{35}$$

- So what can you say about the Kalman filter estimation if the Kalman gain is not an optimal gain?
- Next step is to find out value of $P_{k+1}^-$ where it is needed to seek the value of $e_{k+1}^-$, but it can be defined as when $t_{k+1}$ in the following way,

$$e_{k+1}^- = x_{k+1} - \bar{x}_{k+1}^- \tag{36}$$

where $\bar{x}_{k+1}^-$ can be derived as bellow by assuming $w_k$ has zero mean because it is defined as white noise.

- $$\bar{x}_{k+1}^{-} = \Phi_k \bar{x}_k \tag{37}$$

- $$\bar{x}^{-}_{k+1} = \Phi_k \bar{x}_k \tag{37}$$
- By using Eq. 21,36 and 37, $e^{-}_{k+1}$ can be found as

$$e^{-}_{k+1} = \Phi_k e_k + w_k \tag{38}$$

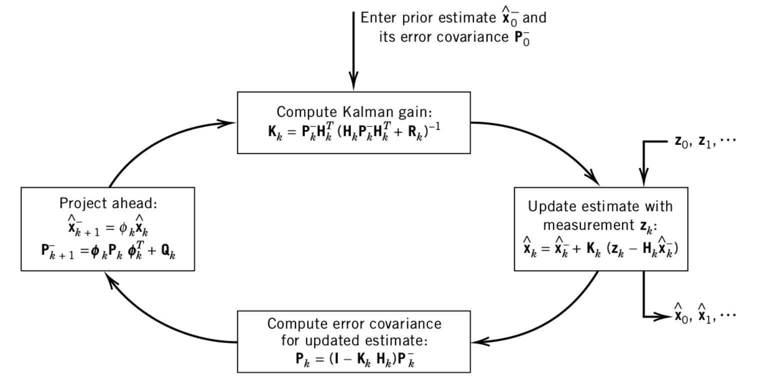
- $$\bar{x}^{-}_{k+1} = \Phi_k \bar{x}_k \qquad (37)$$
- By using Eq. 21,36 and 37, $e^{-}_{k+1}$ can be found as

$$e^{-}_{k+1} = \Phi_k e_k + w_k \qquad (38)$$

- Along with this information, now it is able to define $P^{-}_{k+1}$ as bellow,

$$P^{-}_{k+1} = E[e^{-}_{k+1} {e^{-}_{k+1}}^T] = \Phi_k P_k \Phi_k^T + Q_k \qquad (39)$$

To sum up, let's summarize what we discussed as bellow,



## Example 03

Let's assume we want to track the position of a robot which goes at a constant speed and robot is capable of measuring position x and y from its integrated sensors. Thus, initial step would be to design the model of the robot. Process state vector ($x_k$) can be defined as $[x \; \dot{x} \; y \; \dot{y}]^T$. How can we design the state transition matrix ($\Phi_k$)?

## Example 03

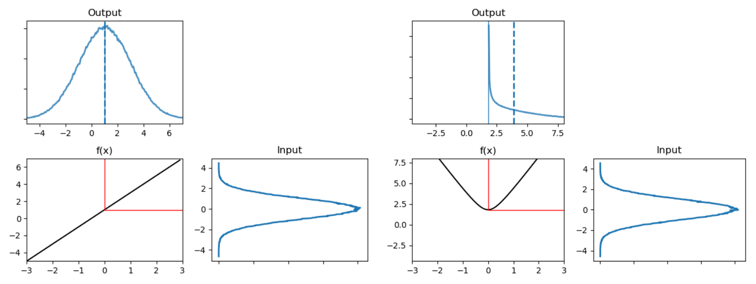
$$\begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix}$$

Can you design R and P by yourself? Next challenging task is to design the matrix $H_k$. When we design this it is needed to go from the state variables to the measurements using the equation $z_k = H_k x_k$. Since robot is capable of measuring its position z vector should be $[x\,y]^T$. So can you guess the dimension of matrix $H_k$? Finally what can you say about the initial conditions? It is needed to decide initial position of the robot, velocity, P and R.

## Example 04

Let's consider the example 03 where to measure the position of the robot instead of one sensor, now it supports 3 sensors: $S_{x1}, S_{x2}$ on the x-direction and $S_{y1}$ on the y-direction, correspondingly. One sensor ($S_{x1}$) gives wheel reports not a position but the number of rotations of the wheels, where 1 revolution yields 1.5 meters of travel.
Can you design the matrix H in this scenario?

The Kalman filter we have studied suites for solving linear equations that are formed as $Ax = b$. What happens if there exists non-linearity in the process of the model?




Passing Gaussian through f(x) = 2x+1 and f(x) = $4\sqrt{x^2 + 0.2}$

- In order to solve this problem, initially Extended Kalman Filter (EKF) was invented shortly after publishing the Kalman filter paper. EKF, itself has some disadvantages which you are going to learn soon.

- In order to solve this problem, initially Extended Kalman Filter (EKF) was invented shortly after publishing the Kalman filter paper. EKF, itself has some disadvantages which you are going to learn soon.
- To overcome those issues, a series of Monte Carlo techniques were proposed throughout the past few decades.

- In order to solve this problem, initially Extended Kalman Filter (EKF) was invented shortly after publishing the Kalman filter paper. EKF, itself has some disadvantages which you are going to learn soon.
- To overcome those issues, a series of Monte Carlo techniques were proposed throughout the past few decades.
- Nonlinearity is handled in EKF by **linearizing** the system at the point of the current estimate and then use a linear Kalman filter to solve the problem while the assuming system is **linear** and has a **unimodal distribution**.

Comparison between KF and EKF for defining process to be estimated and associated measurement relationship where f and h are known functions.

| Name | Process to be estimated | Associated measurement |
|---|---|---|
| KF | $x_{k+1} = \Phi_k x_k + w_k$ | $z_k = H_k x_k + v_k$ |
| EKF | $x_{k+1} = f(x_k, t) + w_k$ | $z_k = h(x_k, t) + v_k$ |

If the system is linear, system should satisfy two fundamental properties: **superposition**, **homogeneity**. Let's consider the two examples to verify whether the considered system is linear or non-linear.

$$\begin{aligned} y = f(x) &= 2x \\ y_1 = 2x_1,\ y_2 &= 2x_2 \\ y_3 &= 2 \cdot (x_1 + x_2) \\ y_1 + y_2 = 2x_1 + 2x_2 &= y_3 \end{aligned} \tag{40}$$

$$\begin{aligned} y = f(x) &= 4\sqrt{x^2 + 0.2} \\ y_1 = 4\sqrt{x_1^2 + 0.2},\ y_2 &= 4\sqrt{x_2^2 + 0.2} \\ y_3 &= 4\sqrt{(x_1 + x_2)^2 + 0.2} \\ y_1 + y_2 = 4\sqrt{x_1^2 + 0.2} + 4\sqrt{x_2^2 + 0.2} &\neq y_3 \end{aligned} \tag{41}$$

- $$y \approx f(x_0) + \frac{df}{dx}|_{x_0} \frac{x - x_0}{1!} + \frac{d^2 f}{dx^2}|_{x_0} \frac{(x - x_0)^2}{2!} + \ldots + \tag{42}$$

  Consider a system is given by $y = f(x) = x^2$, apply Taylor series approximation for $x_0 = 2$ when x equals to 2,2.5 and 3.

# Taylor Series Expansion

- $$y \approx f(x_0) + \frac{df}{dx}|_{x_0} \frac{x - x_0}{1!} + \frac{d^2 f}{dx^2}|_{x_0} \frac{(x - x_0)^2}{2!} + \ldots + \qquad (42)$$
  Consider a system is given by $y = f(x) = x^2$, apply Taylor series approximation for $x_0 = 2$ when x equals to 2,2.5 and 3.
- $$\begin{aligned} y \approx f(x_0) + 2x_0(x - x_0) + \frac{2(x - x_0)^2}{2} + \ldots \\ y = f(x_0) + 2x_0 x - 2x_0^2 \\ y = 4 + 4x - 8 = 4x - 4,\ x_0 = 2 \end{aligned} \qquad (43)$$

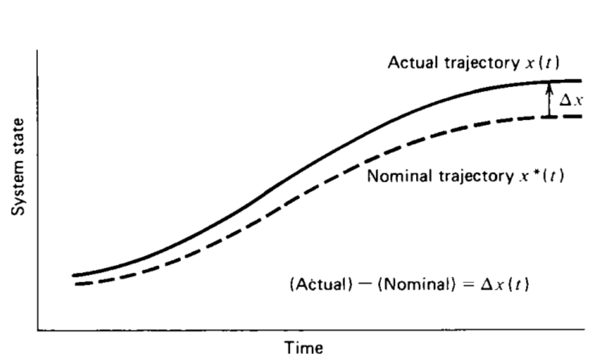



**Figure:** Nominal and actual trajectories for a given function [1]

Let assume approximated trajectory for actual trajectory x(t) as

$$x(t) = x^*(t) + \Delta x(t) \tag{44}$$

then process to be estimated and associated measurement redefine as follow

$$f(x^* + \Delta x, t) + w(t) \approx f(x^*, t) + \left[\frac{\partial f}{\partial x}\right]_{x=x^*} .\Delta x + w(t) \tag{45}$$

$$h(x^* + \Delta x, t) + v(t) \approx h(x^*, t) + \left[\frac{\partial h}{\partial x}\right]_{x=x^*} .\Delta x + v(t) \tag{46}$$

where $\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & . & . & . \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & & & \\ . & & & & \\ . & & & & \\ . & & & & \end{bmatrix}$ and $\frac{\partial h}{\partial x} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & . & . & . \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & & & \\ . & & & & \\ . & & & & \\ . & & & & \end{bmatrix}$

# Comparison between KF and EKF

| KF | EKF |
|---|---|
| | $\boxed{\Phi_k = \left.\frac{\partial f(\mathbf{x}_k, t)}{\partial \mathbf{x}}\right\|_{\mathbf{x}_k}}$ |
| $\hat{\mathbf{x}}_k^- = \Phi_k \mathbf{x}_k$ | $\boxed{\hat{\mathbf{x}}_k^- = f(\mathbf{x}_k, \mathbf{t})}$ |
| $\mathbf{P}_k^- = \Phi_k \mathbf{P}_k \Phi_k^T + \mathbf{Q}_k$ | $\mathbf{P}_k^- = \Phi_k \mathbf{P}_k \Phi_k^T + \mathbf{Q}_k$ |
| | $\boxed{\mathbf{H} = \left.\frac{\partial h(\hat{\mathbf{x}}_k^-)}{\partial \hat{\mathbf{x}}}\right\|_{\hat{\mathbf{x}}_k^-}}$ |
| $\mathbf{y} = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-$<br>$\mathbf{K_k} = \mathbf{P_k^-} \mathbf{H_k}^T (\mathbf{H_k} \mathbf{P_k^-} \mathbf{H_k}^T + \mathbf{R}_k)^{-1}$<br>$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K_k} \mathbf{y}$<br>$\mathbf{P}_k = (\mathbf{I} - \mathbf{K_k} \mathbf{H_k}) \mathbf{P_k^-}$ | $\mathbf{y} = \mathbf{z}_k - \boxed{h(\hat{x}_k^-)}$<br>$\mathbf{K_k} = \mathbf{P_k^-} \mathbf{H_k}^T (\mathbf{H_k} \mathbf{P_k^-} \mathbf{H_k}^T + \mathbf{R}_k)^{-1}$<br>$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K_k} \mathbf{y}$<br>$\mathbf{P}_k = (\mathbf{I} - \mathbf{K_k} \mathbf{H_k}) \mathbf{P_k^-}$ |

## Example 05

Let's assume the relationship between the current and successive pose of the robot is given as follow,

$$\begin{aligned}
\beta &= \frac{d}{w}\tan(\alpha) \\
x &= x - R\sin(\theta) + R\sin(\theta + \beta) \\
y &= y + R\cos(\theta) - R\cos(\theta + \beta) \\
\theta &= \theta + \beta
\end{aligned}$$

where position of the robot is denoted with x and y and heading is denoted with $\theta$.

## Example 05

- Design the state variables vector for this robot ($x_k$)?
- Design the system model ($\Phi_k$) and show that it is given as

$$\breve{}_{\mathbf{k}} = \begin{bmatrix} 1 & 0 & -R\cos(\theta) + R\cos(\theta+\beta) \\ 0 & 1 & -R\sin(\theta) + R\sin(\theta+\beta) \\ 0 & 0 & 1 \end{bmatrix}$$

## Example 05

- A sensor that attached to the robot gives the distances to each visible landmark. Hence, range can be obtained to each sensor as follow. Assume, $p_x^i$ and $p_y^i$ are the distances on x and y direction respectively. $r = \sqrt{(p_x^i - x)^2 + (p_y^i - y)^2}$ where i depicts $i^{th}$ landmark. Relative orientation to each landmark $\phi = \arctan(\frac{p_y^i - y}{p_x^i - x}) - \theta$. Using this information try to obtain the measurement model $(h(\bar{x}_k^-))$?
- The matrix H can be derived as

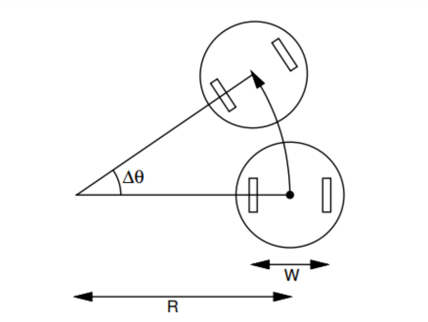
$$\begin{bmatrix} \frac{-p_x+x}{\sqrt{(p_x-x)^2+(p_y-y)^2}} & \frac{-p_y+y}{\sqrt{(p_x-x)^2+(p_y-y)^2}} & 0 \\ -\frac{-p_y+y}{(p_x-x)^2+(p_y-y)^2} & -\frac{p_x-x}{(p_x-x)^2+(p_y-y)^2} & -1 \end{bmatrix}$$

## Example 05

- Is there any information still missing for implementing EKF?
- Let’s try to implement EKF

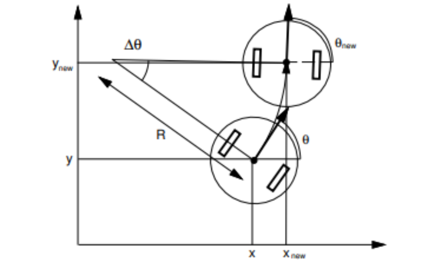
## Example 06

Consider a diff drive robot where $v_l$ and $v_r$ be the left and right wheel velocities, and $w_l$ and $w_r$ are the angular velocities of the left and right wheels, respectively. Assume both the left and right wheel radius is $r$.



## Example 06

Consider a periods of constant motion $\Delta t$, where the robot moves along a circular arc through angle $\Delta\theta$. If both wheel arcs subtend the same angle $\Delta\theta$



# Motion model

- State vector of the diff drive robot

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ y_t \\ \theta_t \end{bmatrix}$$

- Control input

$$\mathbf{u}_t = \begin{bmatrix} \Delta s_{r,t} \\ \Delta s_{l,t} \end{bmatrix}$$

- The motion model

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t) = \begin{bmatrix} x_t \\ y_t \\ \theta_t \end{bmatrix} + \begin{bmatrix} \Delta s_t cos(\theta_t + \Delta\theta_t/2) \\ \Delta s_t sin(\theta_t + \Delta\theta_t/2) \\ \Delta\theta_t \end{bmatrix},$$

where $\Delta s_t = (\Delta s_{r,t} + \Delta s_{l,t})/2$ and $\Delta\theta_t = (\Delta s_{r,t} - \Delta s_{l,t})/2$

# Prediction step

- Covariance estimation, the propagation error can be estimated incorporating control inputs in the following form

$$\left.\begin{array}{r} x \sim N(\mu, \Sigma) \\ y = Ax + B \end{array}\right\} \Rightarrow y \sim N(A\mu + B, A\Sigma A^\top)$$

- Since motion model is non-linear, the first order approximation is considered

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t) \approx A\mathbf{x}_t + B\mathbf{u}_t,$$

where $A = [\frac{\partial f}{\partial x_t} \ \frac{\partial f}{\partial y_t} \ \frac{\partial f}{\partial \theta_t}]$ and $B = [\frac{\partial f}{\partial \Delta s_{r,t}} \ \frac{\partial f}{\partial \Delta s_{l,t}}]$

- Covariance from the prediction step

$$P_t = AP_{t-1}A^\top + BQB^\top,$$

where $Q = \begin{bmatrix} k\,|\Delta s_{r,t}| & 0 \\ 0 & k\,|\Delta s_{r,t}| \end{bmatrix}$

### Example 06

- Assume GPS sensor that is attached to the robot gives current position of the robot. How do you design the H matrix?

# References


ROBERTG BROWN and PATRICKYC HWANG.
**Introduction to random signals and applied kalman filtering(book).**
*New York, John Wiley & Sons, Inc., 1992.* 512, 1992.

HTTP://INVERSEPROBABILITY.COM/TALKS/NOTES/BAYESIAN REGRESSION.HTML.
**Bayesian Regression, 2022.**

**HTTPS://JANAKIEV.COM/BLOG/COVARIANCE MATRIX/.**
**Understanding the Covariance Matrix parametric thoughts, 2022.**

**HTTP://WWW2.IMM.DTU.DK/PUBDB/EDOC/IMM3274.PDF.**
**The Matrix Cookbook kaare brandt petersen, 2022.**

**Gregor Klancar, Andrej Zdesar, Saso Blazic, and Igor Skrjanc.**
***Wheeled mobile robotics: from fundamentals towards autonomous systems.***
Butterworth-Heinemann, 2017.

Roland Siegwart, Illah Reza Nourbakhsh, and Davide Scaramuzza.