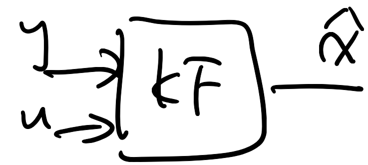
Fall 2021 ME424 Modern Control and Estimation

Lecture Note 7: Kalman Filter - Derivations and Algorithm

Prof. Wei Zhang
Department of Mechanical and Energy Engineering
SUSTech Institute of Robotics
Southern University of Science and Technology

zhangw3@sustech.edu.cn
https://www.wzhanglab.site/

Kalman Filer Preview:



- Given stochastic linear system described by

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k + w_k \\ y_k = C_k x_k + D_k u_k + v_k \end{cases}$$

- **Kalman filter**: compute the "best" estimate of x_k given input-output data history $\{u_j, y_j\}_{j=0}^{k}$

- **Kalman Filter Solution**: $\boxed{\hat{x}_k = \boldsymbol{E}(\boldsymbol{x}_k | \boldsymbol{y}_0, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k)}$

- **Our goal: in-depth understanding of the assumptions, derivations of Kalman filter**

Outline

- **Minimum Mean Squared Estimation (MMSE)**
- Gaussian Random Vectors
- Kalman Filter Derivations
- Summary and Implementation

Fundamental Theorem of Estimation

$y = H\theta + v$

- Suppose we want to estimate the value of a hidden random vector $X \in \mathrm{R}^n$ based on observations of a related vector $Y \in \mathrm{R}^m$.

- We have to know the relationship between X and Y. Suppose we take probabilistic viewpoint of their relations, namely, $(X, Y) \sim f_{XY}(x, y)$

$f_{XY}(x,y)$: pdf. $\qquad$ $p(i,j)$ pmf

- An estimator $\phi(y)$ is a function that maps each measurement $\mathrm{Y} = y$ to an estimate $\hat{x}$

$$y \longrightarrow \underset{\phi(y)}{\boxed{\text{Estimator}}} \longrightarrow \hat{x} \Rightarrow 25.5^\circ$$

e.g. shenzhen 26°

$\hat{x} = \phi(y) = y - 0.5$

$\phi(y) = 2y$?

$= 52^\circ$

- Mean-squared error of an estimator: $\boldsymbol{E}\left(||\boldsymbol{\phi}(\boldsymbol{Y}) - \boldsymbol{X}||^2\right)$ $\Leftarrow$

$(X, Y) \sim f_{XY}(x,y)$

x_1, y_1

x_2, y_2

$\vdots \quad \vdots$

$$= \sum_i \underbrace{||\phi(y_i) - x_i||^2}\cdot \mathrm{Prob}(X = x_i, Y = y_i)$$

- **Example**: Given X,Y joint distribution, compute the mean-squared error for the estimator:

$$\phi(y) \triangleq 2y$$

$$y \to \boxed{} \to 2y$$

		X	
		2	3
Y	1	0.4	0.1
	2	0.2	0.3

$$E(\|\phi(Y)-X\|^2)$$

$$= (\phi(1)-2)^2\cdot 0.4 + (\phi(1)-3)^2\cdot 0.1 + (\phi(2)-2)^2\cdot 0.2 + (\phi(2)-3)^2\cdot 0.3$$

$$= 0 + 0.1 + 0.8 + 0.3 = 1.2$$

By result in next slide: we know $\phi(y) = E(X|Y=y)$ is optimal

check: $\phi(y) = E(X|Y=y) \Rightarrow \begin{cases} \text{If } y=1, \Rightarrow E(X|Y=1) = 2\times 0.8 + 3\times 0.2 = 2.2 \\ \text{If } y=2 \Rightarrow E(X|Y=2) = 2\times 0.4 + 3\times 0.6 = 2.6 \end{cases}$

$$\Rightarrow \phi_{MMSE}(y) = \begin{cases} 2.2, & \text{if } y=1 \\ 2.6 & \text{if } y=2 \end{cases}$$

$$E(\|\phi_{MMSE}(Y)-X\|^2)$$

$$= (2.2-2)^2\cdot 0.4 + (2.2-3)^2\cdot 0.1 + (2.6-2)^2\cdot 0.2 + (2.6-3)^2\cdot 0.3$$

$(X|Y=y) \sim f_{X|Y}(x|y) \Rightarrow$ $=0.2$

- **Theorem: The Minimum Mean-Squared Estimator for $\boldsymbol{X}$ given $\boldsymbol{Y=y}$, that minimizes $\boldsymbol{E\left(||\phi(Y)-X||^2\right)}$ is given by**

$$\hat{X}_{MMSE} = \phi_{MMSE}(y) = E(X|Y=y)$$

Proof : $X \in R^n, Y \in R^m, \phi: \mathrm{R}^m \to R^n$, need to solve $\min_{\phi(\cdot)} E\left(||\phi(Y)-X||^2\right)$

Note: $E\left(||\phi(Y)-X||^2\right) = \int E\left(||\phi(Y)-X||^2 | \, Y=y\right) f_Y(y)dy$, thus we just need to find the estimator $\phi(\cdot)$ to minimize $E\left(||\phi(Y)-X||^2 | \, Y=y\right)$ for each y

$$E\left(||\phi(Y)-X||^2 | Y=y\right) = E\left((\phi(Y)-X)^T(\phi(Y)-X)|Y=y\right)$$
$$= E(\phi(Y)^T\phi(Y) - \phi(Y)^T X - X^T\phi(Y) + X^T X|Y=y)$$
$$= E(\phi(Y)^T\phi(Y)|Y=y) - E(\phi(Y)^T X|Y=y) - E(X^T\phi(Y)|Y=y) + E(X^T X|Y=y)$$
$$= \phi(y)^T\phi(y) - 2\phi(y)^T E(X|Y=y) + E(X^T X|Y=y)$$
$$=\left(\phi(y) - E(X|Y=y)\right)^T\left(\phi(y) - E(X|Y=y)\right) - E(X|Y=y)^T E(X|Y=y) + E(X^T X|Y=y)$$

⇨ Optimal ϕ is thus given by: $\phi(y) = E(X|Y=y)$

- **Remarks:**
  - 1. The MMSE is just the conditional mean !!
  - 2. To compute the MMSE, the general way is to compute the conditional mean directly by definition

$$E(X \mid Y=y) = \int x \cdot f_{X|Y}(x|y)\, dx$$

$\Rightarrow$ very hard to compute in high-dim

  - 3. **Important special case:**

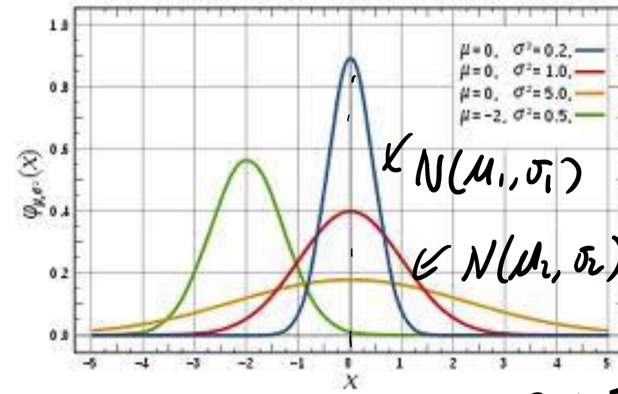
$\in \mathbb{R}^n$ $\in \mathbb{R}^m$

If (X, Y) are jointly Gaussian random vectors, then there is a simple analytical form for the conditional mean

Outline

- Minimum Mean Squared Estimation (MMSE)
- **Gaussian Random Vectors**
- Kalman Filter Derivations
- Summary and Implementation

- **Gaussian Random Vectors**
  - Important due to central limit theorem
  - 1D Gaussian: $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}), \mu \in R, \sigma \in R_+$



$N(\mu_1, \sigma_1)$

$N(\mu_2, \sigma_2)$

$\sigma_2 > \sigma_1$

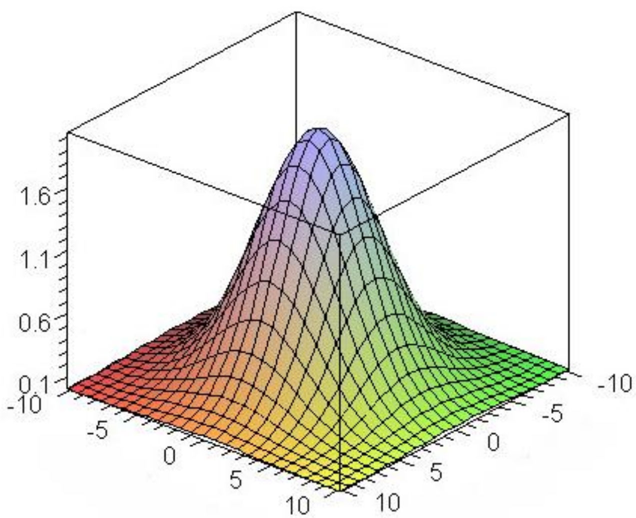
pdf : $$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

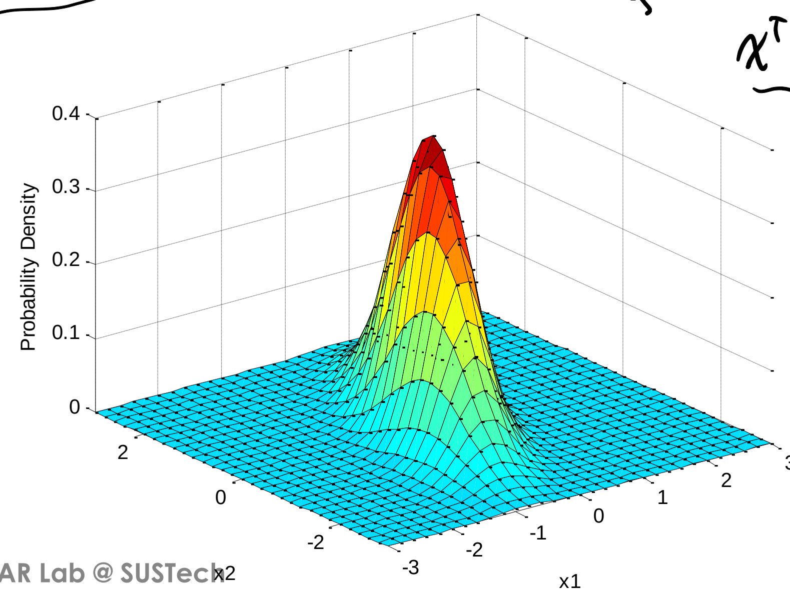
$\int f_X(x)dx = 1$

  - n-D Gaussian: $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \mu \in R^n, \Sigma \in R^{n \times n}$

pdf: $$\boxed{f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(\Sigma))^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}}$$

$x^T P x \le 1$





- Gaussian random vectors have nice properties
  - It can be presented by two parameters:

mean vector and covariance matrix

- **Fact 1: Uncorrelated jointly Gaussian vectors are independent**

$\mathrm{cov}(X,Y)=0 \Rightarrow X \perp Y, \quad f_{XY}(x,y) = f_X(x) f_Y(y)$

- **Fact 2: Linear transformation of Gaussian random vectors are Gaussian**

$X \in N(\mu_X, \Sigma_X) \Rightarrow G = 2X+3, \Rightarrow G$ is Gaussian

- **Fact 3: Conditional Gaussian is Gaussian**
  - If general, checking whether a random variable is Gaussian or not requires computing its probability density function to see whether it is of the form of Gaussian. This can be quite involved.

- **Fact 1: Independence between two Gaussians:**
  - If $X \in R^n, Y \in R^m$ are $\boxed{\text{jointly Gaussians}}$, then $X \perp Y$ if and only if

$$E(XY^T) = E(X) \cdot E(Y)^T$$

$$Cov(X, Y) = 0$$

$\underbrace{\text{Jointly Gaussian}} \triangleq (X, Y)$ as a whole vector $\begin{bmatrix} X \\ Y \end{bmatrix} \in \mathbb{R}^{n+m}$ is Gaussian

$\triangleq f_{XY}(x, y)$ is Gaussian $\iff \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$

(optional)

  - However, if X, Y are both Gaussians, but are not jointly Gaussian, then the above tests do not hold in general
    - See supplemental note on joint Gaussian random vectors

$\boxed{\text{Fact 1.1}}$: if X, Y are both Gaussian & $X \perp Y$

$\Rightarrow \begin{bmatrix} X \\ Y \end{bmatrix} \sim$ Gaussian

$\dot{X_i}$

- **Fact 2: Affine transformation of Gaussian is still Gaussian**

Let $X \sim N(\mu, \Sigma), \mu \in R^n, \Sigma \in R^{n\times n}, A \in R^{m\times n}, b \in R^m$, then

deterministic

$$Z = \underbrace{AX + b}_{} \sim N(\underbrace{A\mu + b}_{\mu_Z}, \underbrace{A\Sigma A^T}_{})$$

$\mu_Z = E(AX+b) = A\mu + b$

$cov(Z) = cov(AX+b, AX+b)$
$= A\Sigma A^T$

- This can be used to test whether a random variable is Gaussian or not

Example: $X \sim N(\mu, \Sigma), X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 1 & 3 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix},$

- Is X_2 Guassian? : ① by definition $f_{X_2}(x_2) = \int_{x_1}\int_{x_3} f(x_1, x_2, x_3)\, dx_1\, dx_3$

② $X_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + 0 \Rightarrow$ affine transformation of X

$\Rightarrow$ Gaussian ✓

- Is $Z = a_2X_2 + a_3X_3$ a Gaussian?

$$Z = \begin{bmatrix} 0 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \Rightarrow \text{Gaussian}$$

- Is $Y = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ a Gaussian?

$$Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} X \Rightarrow \text{Gaussian}$$

- Is $X_2 \perp X_1$? How about X_1 and X_3

$$\mathrm{cov}(X_2, X_1) = 0 \Rightarrow X_2 \perp X_1$$

$$\mathrm{cov}(X_1, X_3) = -1 \Rightarrow \text{not indep.}$$

- **Fact 3: Conditional Gaussian is Gaussian**: Let $X \in R^n, Y \in R^m$ be **jointly** Gaussian with mean μ_X, μ_Y, covariance $\Sigma_X, \Sigma_Y, \Sigma_{XY}, \Sigma_{YX}$, i.e,

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}\right)$$

e.g. $\begin{array}{c} X \\ \\ Y \end{array} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \begin{array}{l} \leftarrow \text{shenzhen} \\ \leftarrow \text{Guangzhou} \\ \leftarrow \text{Beijing} \end{array} \sim N($

then the conditional distribution of X given $Y = y$ is Gaussian

$$X|Y = y \sim N(\mu_{X|Y=y}, \Sigma_{X|Y=y}),$$

where

$$\mu_{X|Y=y} = \underbrace{\mu_X}_{E(X)} + \Sigma_{XY}\Sigma_Y^{-1}(\overset{30}{y} - \underbrace{\mu_Y}_{\overset{25}{E(Y)}})$$

$$\boxed{\Sigma_{X|Y=y} = \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}}$$

how noisy y measurement is

surprise term

(X, Y) jointly Gaussian $\sim f_{XY}(x, y)$ $\Rightarrow$ Given $Y = y$, X is still random its density is $f_{X|Y}(x|y)$, we can show $f_{X|Y}(x|y)$ is a Gaussian

$$\frac{f_{XY}(x,y)}{f_Y(y)}$$

$\mathcal{N}(z_1, \begin{bmatrix} z_2 \\ z_3 \end{bmatrix})$

- **Example**: suppose $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$, where $X \in R^1$, $Y \in R^2$, and $Z \sim N\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 8 \end{bmatrix}\right)$

$\mathrm{cov}(Y, Y)$

: compute the MMSE of X given $Y = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \Rightarrow E(X | Y = \begin{bmatrix} -1 \\ 2 \end{bmatrix})$

1°: (X, Y) jointly Gaussian $\Rightarrow$ $X | Y = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ is Gaussian

2°: $\mu_{X|Y=y} = \mu_X + \Sigma_{XY} \Sigma_Y^{-1} (y - \mu_Y) = 1 + \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 1 \\ 1 & 8 \end{bmatrix}^{-1} \left(\begin{bmatrix} -1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right)$

$=$ some # a

. $\cancel{\Sigma}$ we can find the exact conditional density $f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{\sigma^2}}$

$\Sigma_{X|Y=y} = \Sigma_X - \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX} = 2 - \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 8 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} =$ some # b

- Another example: $Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \in \mathbb{R}^3$, $Z \sim N\left(\begin{bmatrix} -1 \\ 5 \\ 4 \end{bmatrix}, \underbrace{\begin{bmatrix} 1 & -1 & 2 \\ -1 & 4 & 0.5 \\ 2 & 0.5 & 9 \end{bmatrix}}\right)$

$cov(Z_2, Z_3)$

Let $P = \begin{bmatrix} Z_1 \\ Z_3 \end{bmatrix}$, $Q = Z_2$ Find $\underbrace{E(P \mid Q=3)}$

① P, Q jointly Gaussian ($\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_3 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} Z$)

$P \mid Q=3$

$$\underbrace{E(P \mid Q=3)} = E(P) + \underbrace{cov(P,Q)}\,\underbrace{cov(Q)}^{-1}\,(3 - \underbrace{E(Q)}_{5})$$

$$= \begin{bmatrix} -1 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 0.5 \end{bmatrix} \frac{1}{4} (-2) = \begin{bmatrix} -\frac{1}{2} \\ \frac{15}{4} \end{bmatrix}$$

$$cov(P,Q) = \begin{bmatrix} cov(Z_1, Z_2) \\ cov(Z_3, Z_2) \end{bmatrix} = \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}$$

- MMSE example: $X \sim N\left(\begin{bmatrix}4\\4\end{bmatrix}, \begin{bmatrix}4 & 1\\1 & 2\end{bmatrix}\right)$, let $Y = \overbrace{\begin{bmatrix}1 & 0\\1 & 4\end{bmatrix}}^{A} X + V$, where $V \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$, V is independent of X. Find the MMSE of X given $Y = \begin{bmatrix}1\\1\end{bmatrix}$

① we know MMSE of X given $Y=\begin{bmatrix}1\\1\end{bmatrix}$, is $E\left(X \mid Y=\begin{bmatrix}1\\1\end{bmatrix}\right)$

② · X, Y are jointly Gaussian?

$$\left.\begin{array}{l} X \sim N(\;) \\ V \sim N(\;) \\ X \perp V \end{array}\right\} \Rightarrow \begin{bmatrix}X\\V\end{bmatrix} \text{ is Gaussian}$$

$$\Rightarrow \begin{bmatrix}X\\Y\end{bmatrix} = \begin{bmatrix}I & 0\\A & I\end{bmatrix}\begin{bmatrix}X\\V\end{bmatrix} \Rightarrow (X, Y) \text{ jointly Gaussian}$$

③ $\hat{X}_{MMSE} = E\left(X \mid Y=\begin{bmatrix}1\\1\end{bmatrix}\right) = E(X) + \operatorname{cov}(X, Y)\operatorname{cov}(Y)^{-1}\left(\begin{bmatrix}1\\1\end{bmatrix} - E(Y)\right)$

- solution (continue)

3.1 $E(X) = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$.

3.2. $\mathrm{cov}(X,Y) = \mathrm{cov}(X, AX+V)$

$= \mathrm{cov}(X, AX) + \underbrace{\mathrm{cov}(X, V)}$

$= \underbrace{\mathrm{cov}(X,X)} A^T + 0$

$= \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 1 & 9 \end{bmatrix}$

3.3. $\mathrm{cov}(Y)$

$= \mathrm{cov}(AX+V, AX+V)$

$= \mathrm{cov}(AX, AX) + \mathrm{cov}(AX, V) + \underline{\mathrm{cov}(V, AX)} + \underbrace{\mathrm{cov}(V, V)}$

$= A \,\mathrm{cov} X A^T + A \cdot 0 + 0 + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} x & x \\ x & x \end{bmatrix}$

3.4: $E(Y)$

$= AE(X) + E(V)$

$= \begin{bmatrix} 4 \\ 20 \end{bmatrix}$

$\Rightarrow \hat{X}_{MMSE} = \begin{bmatrix} 0.8944 \\ 0.1136 \end{bmatrix}$

Outline

- Minimum Mean Squared Estimation (MMSE)
- Gaussian Random Vectors
- **Kalman Filter Derivations**
- Summary and Implementation

Kalman Filter

- Consider a stochastic linear system described by

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k + w_k \\ y_k = C_k x_k + D_k u_k + v_k \in \mathbb{R}^p \end{cases}$$

→ n-dim Gaussian

1-dim Gaussian

  - $x_k \in R^n$--- system state at time k
  - $y_k \in R^{\cancel{n}p}$--- measurement vector at time k
  - $Y_k \triangleq \left[y_0^T \;\; y_1^T \;\; \dots \;\; y_k^T\right]^T$ --- collection of measurements up to time k
  - $u_k \in \mathrm{R}^{\cancel{p}m}$--- system input at time k (deterministic input)
  - $w_k \in R^n$--- process noise $\sim N(0, Q_k)$
  - $v_k \in R^p$--- measurement noise $\sim N(0, R_k)$
  - Assume $x_0 \sim N(\mu_0, \Phi_0)$, $x_0 \perp w_k, x_0 \perp v_k, w_k \perp v_k, \forall k$

$\left\{ A_k, B_k, C_k, D_k \right\} \to u_k$ are deterministic

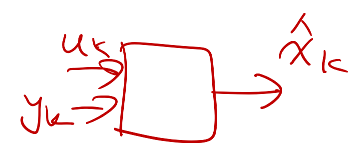
x_1 is Gaussian ? $x_1 = A_0 x_0 + B_0 u_0 + w_0$

$\left.\begin{matrix} x_0 \sim N \\ w_0 \sim N \\ x_0 \perp w_0 \end{matrix}\right\} \Rightarrow \begin{bmatrix} x_0 \\ w_0 \end{bmatrix}$ Gaussian

$\Rightarrow x_1$ Gaussian

- Implications of the above assumption:
  - x_k is Gaussian and y_k is Gaussian for all $k \geq 0$ (VFY)

Everything is jointly Gaussian



- **State estimation problem**: Find the MMSE of x_k given Y_k

, also given $u_0, \cdots u_{k-1}$

- Solution using Fundamental Theorem: $E(x_k|Y_k)$

- **Kalman filter is just a recursive way to compute the conditional mean as new measurement comes in**

$\hat{x}_{k|k+2} \triangleq E(x_k | Y_{k+2})$

- Define: $\hat{x}_{k|k} \triangleq E(x_k|Y_k)$, $\quad \hat{x}_{k|k-1} \triangleq E(x_k|Y_{k-1})$

$\mathrm{cov}(x_k, x_k | Y_k) \Leftrightarrow P_{k|k} \triangleq E\left(\left(x_k - \hat{x}_{k|k}\right)\left(x_k - \hat{x}_{k|k}\right)^T \middle| Y_k\right)$

$\{y_0, \cdots y_k\}$

$P_{k|k-1} \triangleq E\left(\left(x_k - \hat{x}_{k|k-1}\right)\left(x_k - \hat{x}_{k|k-1}\right)^T \middle| Y_{k-1}\right)$

- Simplified notation: $\hat{x}_k \triangleq \hat{x}_{k|k}$, $\quad P_k = P_{k|k}$

$P_{k|k-1}$

$x_k | Y_k \sim \mathcal{N}(\hat{x}_k, P_k)$

- Before deriving Kalman filter, let's work on an example

Example: $\begin{cases} x_{k+1} = A_k x_k + w_k \\ y_k = C_k x_k \end{cases}$

where $x_0 \sim N(0, , \Sigma_x), w_k \sim N(0, \Sigma_w), A = \Sigma_w = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_x = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, C = [1 \quad 2]$.

Compute $\hat{x}_0 = E(x_0|y_0 = 1)$ and $\hat{x}_1 = E(x_1|y_0 = 1, y_1 = -1)$

① $\hat{x}_0 = E(x_0 | y_0 = 1)$, 1.1 we need to find $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ joint distribution

we know $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \sim N\left(\begin{bmatrix} E(x_0) \\ E(y_0) \end{bmatrix}, \begin{bmatrix} cov(x_0) & cov(x_0, y_0) \\ cov(y_0, x_0) & cov(y_0) \end{bmatrix} \right)$

$E(x_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $E(y_0) = [1 \ 2] \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$

$cov(x_0) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \Leftarrow \Sigma_x$

$cov(x_0, y_0) = cov(x_0, x_0) C^T = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$

$cov(y_0, x_0) = [4 \quad 3]$

$cov(y_0) = C \Sigma_x C^T = 10$

$\hat{x}_0 = E(x_0 | y_0 = 1) = E(x_0) + cov(x_0, y_0) cov(y_0)^{-1} [1 - 0]$

$= \begin{bmatrix} 0.4 \\ 0.3 \end{bmatrix}$

$P_0 = cov(x_0, x_0 | y_0 = 1) = cov(x_0, x_0) - cov(x_0, y_0) cov(y_0)^{-1} cov(y_0, x_0)$

$\stackrel{\text{v.f.z.}}{=} \begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}$

Example continue..

$$Y_1$$

② $\hat{x}_1 = E(x_1 \mid \underbrace{y_0 = 1,\ y_1 = -1}) = E\left(x_1 \,\middle|\, \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$

- Conceptually simple approach (Batch approach):

Find joint distribution first: $\begin{bmatrix} x_1 \\ \hline Y_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} E(x_1) \\ E\left(\begin{bmatrix} y_0 \\ y_1 \end{bmatrix}\right) \end{bmatrix}, \begin{bmatrix} \overset{2\times 2}{\mathrm{cov}(x_1)} & \overset{2\times 2}{\mathrm{cov}(x_1, Y_1)} \\ \underset{2\times 2}{\mathrm{cov}(Y_1, x_1)} & \underset{2\times 2}{\mathrm{cov}(Y_1)} \end{bmatrix} \right)$

$$\mathrm{cov}(Y_1) = \begin{bmatrix} \mathrm{cov}(y_0) & \mathrm{cov}(y_0, y_1) \\ \mathrm{cov}(y_1, y_0) & \mathrm{cov}(y_1) \end{bmatrix}$$

$\Rightarrow E\left(x_1 \,\middle|\, Y_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = E(x_1) + \mathrm{cov}(x_1, Y_1) \cdots$

Not recursive, not scalable. as k increases, the covariance matrix grows

- Recursive approach (KF): After obtaining $y_0 = 1$, (~~still~~ still at time 0)

both x_1, y_1 are random, and they are jointly Gaussian

Example continue… i.e.

$$\underbrace{\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}}_{\in \mathbb{R}^3} \Bigg| y_0 = 1 \sim \mathcal{N}\left(\begin{bmatrix} E(x_1 \mid y_0=1) \\ E(y_1 \mid y_0=1) \end{bmatrix}, \begin{bmatrix} \mathrm{cov}(x_1, x_1 \mid y_0=1) & \mathrm{cov}(x_1, y_1 \mid y_0) \\ \mathrm{cov}(y_1, x_1 \mid y_0=1) & \mathrm{cov}(y_1, y_1 \mid y_0) \end{bmatrix} \right)$$

. If we know everything in the above

we ~~ca~~ claim:

$$E(x_1 \mid y_0=1, y_1=-1) = \underbrace{E(x_1 \mid y_0=1)} + \mathrm{cov}(x_1, y_1 \mid y_0=1)\, \mathrm{cov}(y_1 \mid y_0=1)^{-1} \big(-1 - E(y_1 \mid y_0=1)\big)$$

- let's define $Z = \{x_1 \mid y_0 = 1\}$

$Q = \{y_1 \mid y_0 = 1\}$

we know $\begin{bmatrix} Z \\ Q \end{bmatrix}$ jointly Gaussian $\sim \mathcal{N}\left(\begin{bmatrix} E(Z) \\ E(Q) \end{bmatrix}, \begin{bmatrix} \mathrm{cov}(Z) & \mathrm{cov}(Z,Q) \\ \mathrm{cov}(Q,Z), & \mathrm{cov}(Q) \end{bmatrix} \right)$

$$\Rightarrow E(Z \mid Q=-1) = E(Z) + \mathrm{cov}(Z,Q)\,\mathrm{cov}(Q)^{-1}\big(-1 - E(Q)\big)$$

$E(x_1 \mid y_0=1, y_1=-1)$ $\quad E(x_1 \mid y_0=1) +$

$\rightarrow E(x_1 | y_0=1) \,|\, (y_1=-1 | y_0=1)) = E(x_1 | y_0=1, y_1=-1)$

Derivation of Kalman Filter

- Goal of Kalman Filter: obtain recursive formula

$$\begin{Bmatrix} \hat{x}_k \\ P_k \end{Bmatrix} \longrightarrow \begin{Bmatrix} \hat{x}_{k+1} \\ P_{k+1} \end{Bmatrix}$$

- We can compute $\hat{x}_0, P_0$

$k|k \rightarrow k+1|k+1$

sometimes $\hat{x}_0, P_0$ are given directly

If not, we just need to compute $E(x_0 | y_0)$, $P_0 = Cov(x_0 | y_0)$

- Given $\hat{x}_k, P_k$, how to compute $\hat{x}_{k+1}, P_{k+1}$: divide this recursion into two stages: prediction and measurement update

$$\begin{bmatrix} \hat{x}_k \\ P_k \end{bmatrix} \xrightarrow{\text{prediction}} \begin{bmatrix} \hat{x}_{k+1|k} \\ P_{k+1|k} \end{bmatrix} \xrightarrow{\text{measurement update}} \begin{bmatrix} \hat{x}_{k+1} \\ P_{k+1} \end{bmatrix}$$

$k|k$ $\qquad\qquad$ $k+1|k+1$

- **Step 1: prediction** (try to compute $\hat{x}_{k+1|k}, P_{k+1|k}$ using $\hat{x}_k, P_k$

$$\hat{x}_{k+1|k} = E(x_{k+1} \mid Y_k) = E(\underbrace{Ax_k + Bu_k + w_k \mid Y_k}) = A\underbrace{E(x_k \mid Y_k)}_{\hat{x}_k} + Bu_k + E(w_k \mid Y_k)$$

$$= \boxed{A\hat{x}_k + Bu_k} \cdots ①$$

$w_k \perp Y_k$

$\Downarrow$

$E(w_k \mid Y_k) = E(w_k) = 0$

$$P_{k+1|k} = Cov(x_{k+1}, x_{k+1} \mid Y_k)$$

$$= E\left((x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^T \mid Y_k\right)$$

$$= E\left((Ax_k + \cancel{Bu_k} + w_k - A\hat{x}_k - \cancel{Bu_k})(\cdots)^T \mid Y_k\right)$$

$$= \underbrace{E\left(A(x_k-\hat{x}_k)(x_k-\hat{x}_k)^T A^T \mid Y_k\right)} + E(w_k w_k^T \mid Y_k) + \cancel{E\left(A(x_k-\hat{x}_k) w_k^T \mid Y_k\right)}^{\,0} + \cancel{E\left(w_k (x_k-\hat{x}_k)^T A^T \mid Y_k\right)}^{\,0}$$

$A P_k A^T + Q_k \cdots ②$

$w_k \perp Y_k$

$\Rightarrow E(w_k w_k^T) = Q_k$

- Summary of the **prediction step**:

$$\begin{cases} \hat{x}_{k+1|k} = A_k \hat{x}_k + B_k u_k, & ① \\ P_{k+1|k} = A_k P_k A_k^T + Q_k & ② \end{cases}$$

- Key for KF measurement update step:

  - Assume: $X, Y, Z \sim$ jointly Gaussian $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{bmatrix}, \begin{bmatrix} \ \end{bmatrix} \right)$

- $E(X \mid Y=y, Z=z)$

  "Batch"

  - ① Direct approach: view $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$ as a big vector $E\left(X \,\middle|\, \underline{\begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{pmatrix} y \\ z \end{pmatrix}}\right)$

    using conditional Gaussian.

  - ② Recursive: ⓐ $E(X \mid Y=y)$ upon knowing $Y=y$

    ⓑ Update $\underline{E(X \mid Y=y, Z=z)}$ upon knowing $Z=z$

    using $\underline{E(X \mid Y=y)}$

① ② $\Rightarrow$ the same result

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{pmatrix}, \begin{pmatrix} \mathrm{cov}\,X & \mathrm{cov}\,XY & \mathrm{cov}\,XZ \\ \mathrm{cov}\,YX & \cdots & \\ \vdots & & \end{pmatrix} \right)$$

$$E(X \mid Z=z, Y=y)$$

Event $Y=y$

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim N\left(\begin{bmatrix} \tilde{\mu}_X \\ \tilde{\mu}_Z \end{bmatrix}, \begin{bmatrix} \widetilde{\mathrm{cov}}(X,X), & \widetilde{\mathrm{cov}}(X,Z) \\ \widetilde{\mathrm{cov}}(Z,X), & \widetilde{\mathrm{cov}}(Z) \end{bmatrix} \right)$$

Event $Z=z$

$$X \sim N(\tilde{\tilde{\mu}}_X, \tilde{\widetilde{\mathrm{cov}}}\,X)$$

$$E(X \mid Y=y, Z=z)$$

$$\tilde{\tilde{\mu}}_X = \tilde{\mu}_X + \widetilde{\mathrm{cov}}(X,Z)\,\widetilde{\mathrm{cov}}(Z)^{-1}(z - \tilde{\mu}_Z)$$

$E(X \mid Y=y)$

$\overbrace{\text{additional surprise due}}^{\text{beyond } Y=y}$ $Z=z$

- **Step 2: measurement update**
  - We want $\hat{x}_{k+1} = E(x_{k+1}|Y_{k+1}) = E(x_{k+1}|Y_k, y_{k+1})$
  - Up to now, we have $\boxed{\hat{x}_{k+1|k}}$ and $P_{k+1|k}$, i.e. the mean and covariance of conditional random variable $x_{k+1}|Y_k$
  - How to find the mean and covariance of $x_{k+1}|\{Y_k, y_{k+1}\}$
  - Define $Z = x_{k+1}|Y_k,\ W = y_{k+1}|Y_k \Rightarrow \hat{x}_{k+1} = E(Z|W)$

Approach 1°: batch. Find joint $\begin{pmatrix} x_{k+1} \\ \begin{pmatrix} Y_k \\ y_{k+1} \end{pmatrix} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \ \end{bmatrix}, \begin{bmatrix} \ \end{bmatrix}\right)$

then we can use conditional mean formula to compute $E(x_{k+1}|Y_k, y_{k+1})$

Approach 2°: Recursive: using the previous result $\hat{x}_{k+1|k}$, $P_{k+1|k}$ to compute $\begin{pmatrix} \hat{x}_{k+1} \\ P_{k+1} \end{pmatrix}$

- Derivation (continue)

Find condition joint

$$\left.\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix}\right| \underbrace{Y_k}_{(y_0 \cdots y_k)} \sim N\left(\begin{bmatrix} E(x_{k+1}/Y_k) \\ E(y_{k+1}/Y_k) \end{bmatrix}, \begin{bmatrix} cov(x_{k+1}, x_{k+1}/Y_k) & cov(x_{k+1}, y_{k+1}/Y_k) \\ cov(y_{k+1}, x_{k+1}/Y_k) & cov(y_{k+1}, y_{k+1}/Y_k) \end{bmatrix} \right)$$

$\Rightarrow$ If we define $Z = x_{k+1}/Y_k$, $W = y_{k+1}/Y_k$

$$\begin{pmatrix} Z \\ W \end{pmatrix} \sim N\left(\begin{bmatrix} \quad \end{bmatrix}, \begin{bmatrix} \quad \end{bmatrix} \right)$$

$$E(Z/W = y_{k+1}) = E(Z) + cov(Z, W)\, cov(W)^{-1} (y_{k+1} - E(W))$$

$$= \underbrace{E(x_{k+1}/Y_k)}_{①} + \underbrace{cov(x_{k+1}, y_{k+1}/Y_k)}_{②}\, \underbrace{cov(y_{k+1}, y_{k+1}/Y_k)}_{③}^{-1} (y_{k+1} - \underbrace{E(y_{k+1}/Y_k)}_{④})$$

- ① $E(x_{k+1} | Y_k) = \hat{x}_{k+1|k}$ ← given by prediction step.

④ $E(y_{k+1} | Y_k) = E(Cx_{k+1} + Du_{k+1} + v_{k+1} | Y_k)$

$= C\underbrace{E(x_{k+1} | Y_k)} + Du_{k+1} + 0$

$= C\hat{x}_{k+1|k} + Du_{k+1}$

② $cov(x_{k+1}, y_{k+1} | Y_k) = cov(x_{k+1}, Cx_{k+1} + Du_{k+1} + v_{k+1} | Y_k)$

$= \underbrace{cov(x_{k+1}, x_{k+1} | Y_k)}_{P_{k+1|k}} \cdot C^T + 0 + 0$

$= P_{k+1|k} C^T$

$\boxed{③}$

- Complete derivation will be posted online

Outline

- Minimum Mean Squared Estimation (MMSE)
- Gaussian Random Vectors
- Kalman Filter Derivations
- **Summary and Implementation**

Kalman Filter Notations

System model

$$x_{k+1} = A_k x_k + B_k u_k + w_k$$
$$y_k = C_k x_k + D_k u_k + v_k$$

Noise model

$$w_k \sim N(0, Q_k)$$
$$v_k \sim N(0, R_k)$$

Measurement history

$$Y_k = \{y_0, y_1, \ldots, y_k\}$$

Filtered estimate

$$\hat{x}_k = \hat{x}_{k|k} = E(x_k | Y_k)$$

Filter error

$$e_k = x_k - \hat{x}_k, \text{ with } E(e_k) = 0$$

Filter error covariance

$$P_k = E\left(e_k e_k^T\right) = E\left(e_k e_k^T \mid Y_k\right)$$

Predicted estimate

$$\hat{x}_{k|k-1} = E\left(x_k \middle| Y_{k-1}\right)$$

Prediction error

$$e_{k|k-1} = x_k - \hat{x}_{k|k-1}, \text{ with } E\left(e_{k|k-1}\right) = 0$$

Prediction error covariance

$$P_{k|k-1} = E\left(e_{k|k-1} e_{k|k-1}^T\right) = E\left(e_{k|k-1} e_{k|k-1}^T \middle| Y_{k-1}\right)$$

Some facts:

- Unbiasedness: $E(e_k) = 0 \quad E\left(e_{k|k-1}\right) = 0$
- Error covariance equals the conditional error covariance

$$E\left(e_k e_k^T\right) = E\left(e_k e_k^T \mid Y_k\right) \qquad E\left(e_{k|k-1} e_{k|k-1}^T\right) = E\left(e_{k|k-1} e_{k|k-1}^T \middle| Y_{k-1}\right)$$

- Mean squared error: $E\left(\left\|e_k\right\|^2\right) = \text{trace}\left(P_k\right), \; E\left(\left\|e_{k|k-1}\right\|^2\right) = \text{trace}\left(P_{k|k-1}\right)$

Kalman Filter Diagram

$$\begin{Bmatrix}\hat{x}_{k-1}\\ P_{k-1}\end{Bmatrix} \xrightarrow{\text{prediction}} \begin{Bmatrix}\hat{x}_{k|k-1}\\ P_{k|k-1}\end{Bmatrix} \xrightarrow{\text{Measurement update}} \begin{Bmatrix}\hat{x}_{k}\\ P_{k}\end{Bmatrix} \xrightarrow{\text{prediction}} \begin{Bmatrix}\hat{x}_{k+1|k}\\ P_{k+1|k}\end{Bmatrix} \cdots\cdots$$

$E(x_{0|-1}) = E(x_0)$

$cov(x_{0|-1}) = cov(x_0)$

$\{\hat{x}_{0|-1}, P_{0|-1}\}$ Initial condition

$cov(x_{k+1}, y_{k+1} | Y_k)$

$E(y_{k+1} | Y_k)$

Prediction

1. Predict state ahead

$$\hat{x}_{k|k-1} = A_{k-1}\hat{x}_{k-1} + B_{k-1}u_{k-1}$$

2. Prediction error covariance

$$P_{k|k-1} = A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1}$$

Measurement Update

1. Compute Kalman gain

$$K_k = \underbrace{P_{k|k-1}C_k^T}\left(C_k \underbrace{P_{k|k-1}C_k^T} + R_k\right)^{-1}$$

2. Update estimate

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k\underbrace{\left(y_k - C_k\hat{x}_{k|k-1} - D_k u_k\right)}_{\text{surprise}}$$

Innovation

3. Update error covariance

$$P_k = \left(I - K_k C_k\right)P_{k|k-1}$$

$k = k+1$

Initial condition $\{\hat{x}_0, P_0\}$

$E(x_0 | y_0)$

If the system (A, C) is observable

then we can arbitrary pick $(\hat{x}_0, P_0)$

not necessarily the true

Either of the two initial conditions is enough to start the iteration

- in simulation. e.g. $\underbrace{x = \text{randn}(2,1)}$ $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$ $\sim N(x_1)$

$$V \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, R\right)$$

$$V = R^{\frac{1}{2}} \cdot x$$

→ matrix square root (for positive semidefinite matrix)

Coding Example

$$R^{\frac{1}{2}} \cdot R^{\frac{1}{2}} = R$$

$$\operatorname{cov}(V, V) = R^{\frac{1}{2}} \operatorname{cov}(x) \left(R^{\frac{1}{2}}\right)^T = R \cdot I = R$$