Kalman Filter

Embedded Systems Lecture Notes

Kasım Sinan Yıldırım

Random Variables

- Let X be a random variable with a **probability density function** $p(x)$.
- **Expected value** of X is given by:

$$\mathbb{E}[X] = \mu = \int_{-\infty}^{\infty} xp(x)dx \tag{1}$$

- **Variance** of X is given by:

$$Var[X] = \sigma^2 = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{2}$$

Example

As an example, if X is a random variable of *Gaussian* distribution:

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \implies p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3}$$

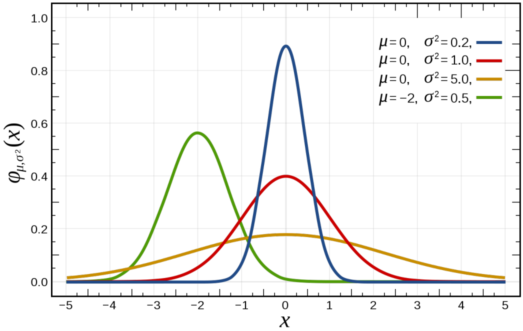

Figure: Gaussian Distribution

Random Variables

- **Covariance** of two random variables X_1 and X_2:

$$\begin{aligned} cov(X_1, X_2) = \sigma_{x_1 x_2} &= \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 - \mu_1)(x_2 - \mu_2)p(x_1, x_2)dx_1 dx_2 \end{aligned} \tag{4}$$

where $p(x_1, x_2)$ is the joint probability.

Random Vectors

- Let $\mathbf{X} = [X_1, \ldots, X_n]^T$ be a vector of random variables. Then

$$\bar{\mathbf{X}} = \boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} X p(x_1, \ldots, x_n) dx_1 \ldots dx_n. \tag{5}$$

- The covariance matrix becomes:

$$\begin{aligned} cov(\mathbf{X}) = \boldsymbol{\Sigma} &= \mathbb{E}[(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T] \\ &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (X - \bar{\mathbf{X}})(X - \bar{\mathbf{X}})^T p(X) dX_1 \ldots dX_n \\ &= \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} & \ldots & \sigma_{x_1 x_n} \\ \sigma_{x_1 x_2} & \sigma_{x_2}^2 & \ldots & \sigma_{x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_1 x_n} & \sigma_{x_2 x_n} & \ldots & \sigma_{x_n}^2 \end{pmatrix} \end{aligned} \tag{6}$$

Gaussian Random Vectors

- A random vector $\mathbf{X} \in \mathbb{R}^n = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is Gaussian if it has density

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \implies p(X) = \frac{1}{\sqrt{2\pi^n \det(\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(X-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(X-\boldsymbol{\mu})} \tag{7}$$

where

$$\bar{\mathbf{X}} = \boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(X) dX_1 \dots dX_n, \tag{8}$$

$$\boldsymbol{\Sigma} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (X - \bar{\mathbf{X}})(X - \bar{\mathbf{X}})^T p(X) dX_1 \dots dX_n \tag{9}$$

Linear Transformation of Random Vectors

- Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\mathbf{X}})$ be Gaussian random vector. Consider linear transformation of $\mathbf{X}$:

$$\mathbf{Z} = \mathbf{A}\mathbf{X} + \mathbf{b} \tag{10}$$

where $\mathbf{A} \in \mathbb{R}^{mxn}$ and $\mathbf{b} \in \mathbb{R}^{m}$.

- $\mathbf{Z}$ is a random Gaussian vector with mean and covariance as below:

$$\begin{aligned}
\mathbb{E}[\mathbf{Z}] &= \mathbb{E}[\mathbf{AX} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b} \\
&= \mathbf{A}\boldsymbol{\mu}_{\boldsymbol{X}} + \mathbf{b}; \qquad (11) \\
\boldsymbol{\Sigma}_{\boldsymbol{Z}} &= \mathbb{E}[(Z - \bar{\mathbf{Z}})(Z - \bar{\mathbf{Z}})^T] = \mathbb{E}[\mathbf{A}(X - \bar{\mathbf{X}})(X - \bar{\mathbf{X}})^T \mathbf{A}^T] \\
&= \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{A}^T \qquad (12)
\end{aligned}$$

Discrete Time Linear Dynamical System

$$x_{t+1} = \mathbf{A}x_t + w_t, \qquad z_t = \mathbf{C}x_t + v_t \tag{13}$$

where

- $x_t \in \mathbb{R}^n$ is the *state*,
- $w_t \in \mathbb{R}^n \sim \mathcal{N}(0, \Sigma_w)$ is the *state noise*,
- $z_t \in \mathbb{R}^p$ is the *observed output*,
- $v_t \in \mathbb{R}^p \sim \mathcal{N}(0, \Sigma_v)$ is the *measurement noise*.

Statistical Properties:

- Sensor noise v_t is independent of x_t
- State noise w_t is independent of x_t and z_t
- Process is *Markov*; i.e. if you know x_{t-1}, then knowledge of $x_{t-2}, \ldots, x_0$ does not give any more information about x_t:

$$x_t | x_0 \ldots x_{t-1} = x_t | x_{t-1} \tag{14}$$

System View

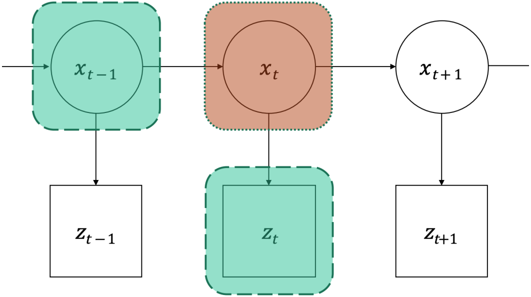

Figure: System View

- x_{t-1} is known and z_t is measured
- Estimate x_t based on knowledge x_{t-1} and z_t

"I measure z_t and I have knowledge to some extent about my previous state x_{t-1}; so that is my current state?"

Bayesian Modeling

We have

$$x_{t+1} = \mathbf{A}x_t + w_t, \qquad z_t = \mathbf{C}x_t + v_t$$

where $w_t \sim \mathcal{N}(0, \Sigma_w)$ and $v_t \sim \mathcal{N}(0, \Sigma_v)$.

- We will **assume** that x_t is a *Gaussian random vector*; i.e.

$$x_t \sim \mathcal{N}(\mu_t, \Sigma_t) \tag{15}$$

- Let's apply linear transformations: we know x_t and we can measure z_t; therefore we can predict x_{t+1}

$$p(x_{t+1}|x_t) = \mathbf{A}p(x_t) + w_t \tag{16}$$

$$p(z_t|x_t) = \mathbf{C}p(x_t) + v_t \tag{17}$$

Bayesian Kalman Filtering

- Bayes' Rule

$$p(\alpha|\beta) = \frac{p(\beta|\alpha)p(\alpha)}{p(\beta)} \tag{18}$$

where $p(\alpha)$ is **prior** distribution, $p(\beta|\alpha)$ is the **likelihood** and $p(\alpha|\beta)$ is **posterior** distribution.

- We have the following distributions:

$$p(x_t|x_{t-1}) \sim \mathcal{N}\left(\mathbf{A}\mu_{t-1}, \mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w\right) \tag{19}$$

$$p(z_t|x_t) \sim \mathcal{N}\left(\mathbf{C}\mu_t, \mathbf{C}\Sigma_t\mathbf{C}^T + \Sigma_v\right) \tag{20}$$

- Apply Bayes' Rule:

$$p(x_t|z_t, x_{t-1}) = \frac{p(z_t|x_t, x_{t-1})p(x_t|x_{t-1})}{p(z_t)} \tag{21}$$

Bayesian Kalman Filtering

- Posterior distribution is also Gaussian. Maximize posterior distribution:

$$\begin{aligned}\hat{x}_t &= \arg\max_{x_t} p(x_t|z_t, x_{t-1}) \\ &= \arg\max_{x_t} \frac{p(z_t|x_t, x_{t-1})p(x_t|x_{t-1})}{p(z_t)} \\ &= \arg\max_{x_t} p(z_t|x_t, x_{t-1})p(x_t|x_{t-1}) \end{aligned} \tag{22}$$

- Therefore, we get

$$\hat{x}_t = \arg\max_{x_t} \mathcal{N}\left(\mathbf{C}\mu_t, \mathbf{C}\Sigma_t\mathbf{C}^T + \Sigma_v\right)\mathcal{N}\left(\mathbf{A}\mu_{t-1}, \mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w\right) \tag{23}$$

Bayesian Kalman Filtering

Let $\mathbf{R} = \mathbf{C}\Sigma_t\mathbf{C}^T + \Sigma_v$ and $\mathbf{P} = \mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w$. Using the **log-likelihood** function, we obtain:

$$\hat{x}_t = \arg\min_{x_t} (z_t - \mathbf{C}x_t)\mathbf{R}^{-1}(z_t - \mathbf{C}x_t)^T + (x_t - \mathbf{A}x_{t-1})\mathbf{P}^{-1}(x_t - \mathbf{A}x_{t-1})^T \tag{24}$$

Take the derivative with respect to x_t and after several steps we reach:

$$\hat{x}_t = \mathbf{A}x_{t-1} + K(z_t - \mathbf{CA}x_{t-1}) \tag{25}$$

where K is the **Kalman Gain** which is defined as:

$$K = \mathbf{PC}^T(\mathbf{R} + \mathbf{CP}^T\mathbf{C}^T)^{-1} \tag{26}$$

Bayesian Kalman Filtering

Therefore, the values of $\boldsymbol{\mu_t}$ and $\boldsymbol{\Sigma_t}$ that maximizes the posterior becomes:

$$\mu_t = \mathbf{A}\mu_{t-1} + K(z_t - \mathbf{CA}\mu_{t-1}); \tag{27}$$

$$\Sigma_t = (\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w) - K\mathbf{C}(\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w). \tag{28}$$

where K is the **Kalman Gain** which is defined as:

$$K = (\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w)\mathbf{C}^T(\mathbf{C}\Sigma_t\mathbf{C}^T + \Sigma_v + \mathbf{C}(\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w)^T\mathbf{C}^T)^{-1} \tag{29}$$

Therefore, our posterior distribution becomes:

$$x_t \sim \mathcal{N}(\mu_t, \Sigma_t) \tag{30}$$

The Algorithm

Kalman Filter Pseudocode

Kalman Filter$(\mu_{t-1}, \Sigma_{t-1}, z_t)$

$K = (\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w)\mathbf{C}^T(\mathbf{C}\Sigma_t\mathbf{C}^T + \Sigma_v + \mathbf{C}(\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w)^T\mathbf{C}^T)^{-1}$

$\mu_t = \mathbf{A}\mu_{t-1} + K(z_t - \mathbf{C}\mathbf{A}\mu_{t-1})$

$\Sigma_t = (\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w) - K\mathbf{C}(\mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Sigma_w)$

Return (μ_t, Σ_t)

- It is an optimal estimator
  - infers parameters of interest from inaccurate and uncertain observations.
- It is **recursive** so that new measurements can be processed as they arrive.