NA 568 - Winter 2026

Kalman Filtering

Maani Ghaffari

January 12, 2026

UNIVERSITY OF MICHIGAN

From Bayes Filter to Kalman Filter

Bayes filter maintains a belief

$$bel(x_k) \;=\; p(x_k \mid z_{1:k}, u_{1:k}).$$

Kalman filtering: Assume the models are *linear* and all uncertainties are *Gaussian*.

$$\begin{aligned} x_k &= F_k x_{k-1} + G_k u_k + w_k, &\qquad w_k &\sim \mathcal{N}(0, Q_k) \\ z_k &= H_k x_k + v_k, &\qquad v_k &\sim \mathcal{N}(0, R_k) \end{aligned}$$

Consequence: the belief stays Gaussian for all time:

$$bel(x_k) \equiv \mathcal{N}(\mu_k, \Sigma_k) \quad \Rightarrow \quad \text{we only track } (\mu_k, \Sigma_k).$$

Today: derive the *predict/update* rules for (μ_k, Σ_k) and interpret the Kalman gain.

Two Gaussian identities (we will use repeatedly)

1) Affine transform: if $X \sim \mathcal{N}(\mu,\Sigma)$ and $Y = AX + b$, then

$$Y \sim \mathcal{N}(A\mu + b,\, A\Sigma A^\top).$$

2) Conditioning (joint Gaussian): If

$$\begin{bmatrix} x \\ z \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_z \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix} \right),$$

then the conditional $p(x \mid z)$ is Gaussian with

$$\mu_{x|z} = \mu_x + \Sigma_{xz}\Sigma_{zz}^{-1}(z - \mu_z), \qquad \Sigma_{x|z} = \underset{11}{\Sigma_{xx}} - \underset{12}{\Sigma_{xz}}\underset{22}{\Sigma_{zz}^{-1}}\underset{21}{\Sigma_{zx}}.$$

Interpretation: "posterior mean = prior mean + (gain)×(innovation)," i.e., a linear update rule.

Recap: Univariate Normal Distribution

The univariate (one-dimensional) *Gaussian (or normal) distribution* with mean μ and variance σ^2 has the following Probability Density Function (PDF).

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$$

We often write $X \sim \mathcal{N}(\mu, \sigma^2)$ or $\mathcal{N}(x; \mu, \sigma^2)$ to imply that X follows a Gaussian distribution with mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathbb{V}[X]$.

The multivariate Gaussian (normal) distribution of an n-dimensional random vector $X \sim \mathcal{N}(\mu, \Sigma)$, with mean $\mu = \mathbb{E}[X]$ and covariance $\Sigma = \mathrm{Cov}[X] = \mathbb{E}[(X-\mu)(X-\mu)^\mathsf{T}]$ is

$$p(x) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu))$$

Let $x = \mathrm{vec}(x_1, x_2)$ and $X \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{bmatrix} 0.0 \\ 0.5 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$$

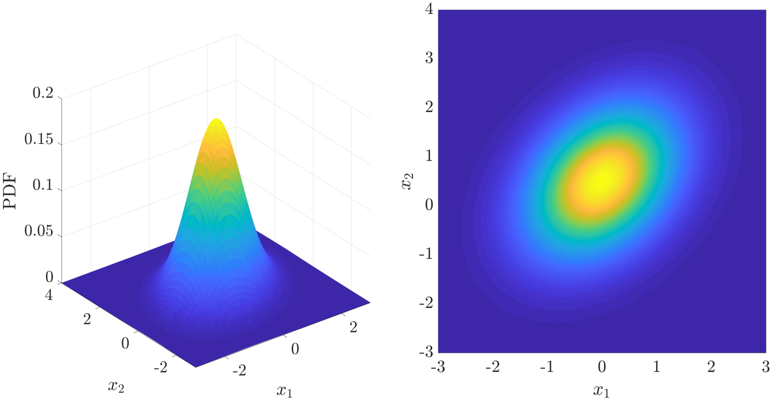


Figure: Left, two-dimensional PDF; right, top view of the first plot.

Marginalization and Conditioning of Normal Distribution

Let X and Y be jointly Gaussian random vectors

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^{\mathsf{T}} & B \end{bmatrix})$$

then the marginal distribution of X is

$$X \sim \mathcal{N}(\mu_x, A)$$

and the conditional distribution of X given Y is

$$X|Y_{=y} \sim \mathcal{N}(\mu_x + CB^{-1}(y - \mu_y), A - CB^{-1}C^{\mathsf{T}})$$

Visualizing multivariate Gaussian

Let $x = \mathrm{vec}(x_1, x_2)$ and $X \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{bmatrix} 0.0 \\ 0.5 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$$

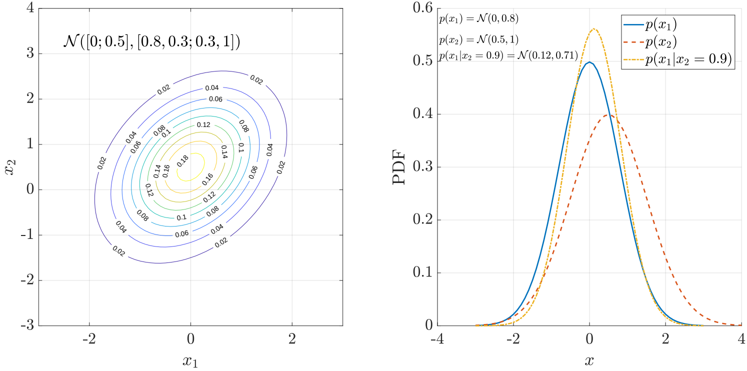


Figure: Left, the contour plot of the PDF; right, the marginals and the conditional distribution of $p(x_1|x_2 = 0.9)$.

A discrete-time random process (random sequence), $\mathbf{w}_k$, is called white noise if:

$$\mathbb{E}[\mathbf{w}_k \mathbf{w}_j^\mathsf{T}] = \mathbf{Q}_k \delta_{kj}$$

白噪声当且仅当同一时刻
有相关系数，不同时刻不相关。

where the Kronecker δ_{kj} is

$$\delta_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}$$

- The state of a dynamic system excited by white noise

  $$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{w}_k)$$

  is a discrete-time Markov process or Markov sequence.

- The state of a linear dynamic system excited by white Gaussian noise

  $$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{w}_k$$

  is called a Gauss-Markov process.

- Assuming the initial condition is Gaussian, because of linearity $\mathbf{x}_k$ is Gaussian and because of the whiteness of the process noise it is Markov.

Kalman Filter (KF) Assumptions

- The state, $\mathbf{x}_k$, evolves according to a known linear dynamic equation with:
- known inputs, $\mathbf{u}_k$;
- an additive process noise, $\mathbf{w}_k$, which is a zero-mean white (uncorrelated) process with known covariance $\mathbf{Q}_k$;

$$\mathbf{x}_k^- = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{G}_k \mathbf{u}_k + \mathbf{w}_k$$

xk-就是xk|k-1

- Measurement model is a known linear function of the state with:
- an additive measurement noise, $\mathbf{v}_k$, which is a zero-mean white (uncorrelated) process with known covariance $\mathbf{R}_k$;

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k^- + \mathbf{v}_k$$

观测应该和xk|k-1这个预测有仿射关系

为什么Kalman滤波是贝叶斯滤波？
因为满足
1.预测步：实现从p(xt|z1:t-1,ut-1)到p(xt|z1:t-1,ut)的转变——添加ut信息
2.更新步：实现从p(xt|z1:t-1,ut)到p(xt|z1:t,ut)的转变——添加zt信息

- Initial state is assumed to be a random variable with known mean (initial estimate) and known covariance (initial uncertainty).

- Initial state and noises are all mutually uncorrelated.

- Initial state $\mathbf{x}_0$ (with possibly given prior information $\mathbf{z}_0$): $\mathbb{E}[\mathbf{x}_0|\mathbf{z}_0] = \hat{\mathbf{x}}_0$ and $\mathrm{Cov}[\mathbf{x}_0|\mathbf{z}_0] = \mathbf{P}_0$ 初始状态的协方差为P0

- Process and measurement noise sequences are white with known covariances:
  $\mathbb{E}[\mathbf{w}_k] = \mathbf{0}$, $\mathbb{E}[\mathbf{w}_k\mathbf{w}_j^\mathsf{T}] = \mathbf{Q}_k\delta_{kj}$, and
  $\mathbb{E}[\mathbf{v}_k] = \mathbf{0}$, $\mathbb{E}[\mathbf{v}_k\mathbf{v}_j^\mathsf{T}] = \mathbf{R}_k\delta_{kj}$ 过程噪声和观测噪声的协方差为Q和R

- All the above are uncorrelated.

- State and measurement prediction, a.k.a., time update;
- State update, a.k.a., correction or measurement update.

协方差理论上应该满足两件事：

1. **对称**：$P = P^T$
2. **半正定**：$P \succeq 0$（不应出现负方差）

Joseph form 的优点是：只要 $P^- \succeq 0$、$R \succeq 0$，那么

$$AP^-A^T \succeq 0, \quad KRK^T \succeq 0$$

两项相加仍然 $\succeq 0$，并且天然更接近对称。

而第 7 行 $P^+ = (I - KH)P^-$：

- 由于舍入误差，它**很容易不对称**（少了右乘 $(I-KH)^T$ 这半边的"对称保护"）
- 也更容易把本来很小的方差"减"成负的（尤其当滤波已经收敛、P^- 很小、或者系统/量测矩阵尺度差异大时）

- $\underline{\theta}$ and $\underline{W}$ are jointly Gaussian:

$$\begin{bmatrix} \underline{\theta} \\ \underline{w} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_\theta \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} R_\theta & 0 \\ 0 & R_w \end{bmatrix}}_{} \right)$$

- Therefore

$$\begin{bmatrix} \underline{X} \\ \underline{\theta} \end{bmatrix} = \underbrace{\begin{bmatrix} A & I_N \\ I_p & 0 \end{bmatrix}}_{} \begin{bmatrix} \underline{\theta} \\ \underline{w} \end{bmatrix}$$

$$\begin{bmatrix} A & I_N \\ I_p & 0 \end{bmatrix} \begin{bmatrix} R_\theta & 0 \\ 0 & R_w \end{bmatrix} \begin{bmatrix} A^T & I_p^T \\ I_N^T & 0^T \end{bmatrix}$$

$$= \begin{bmatrix} AR_\theta & I_N \\ R_\theta & 0 \end{bmatrix}$$

$$\sim N\left(\begin{bmatrix} A\mu_\theta \\ \mu_\theta \end{bmatrix}, \begin{bmatrix} AR_\theta A^T + R_w & AR_\theta \\ R_\theta A^T & R_\theta \end{bmatrix} \right)$$

KF Algorithm

Joseph form:

$$P^+ = (I - KH)P^-(I - KH)^T + KRK^T$$

在 $K = P^-H^T(HP^-H^T + R)^{-1}$ 时，可以严格化简成:

$$P^+ = (I - KH)P^-$$

所以理论上它们算的是同一个 P^+。

第7行和第8行的等价性

Algorithm 1 `Kalman-filter`

Require: belief mean $\boldsymbol{\mu}_{k-1}$, belief covariance $\boldsymbol{\Sigma}_{k-1}$, action $\mathbf{u}_k$, measurement $\mathbf{z}_k$;

1: $\boldsymbol{\mu}_k^- \leftarrow \mathbf{F}_k\boldsymbol{\mu}_{k-1} + \mathbf{G}_k\mathbf{u}_k$ xk|k-1均值 ▷ predicted mean

2: $\boldsymbol{\Sigma}_k^- \leftarrow \mathbf{F}_k\boldsymbol{\Sigma}_{k-1}\mathbf{F}_k^\mathsf{T} + \mathbf{Q}_k$ xk|k-1协方差 ▷ predicted covariance

3: $\boldsymbol{\nu}_k \leftarrow \mathbf{z}_k - \mathbf{H}_k\boldsymbol{\mu}_k^-$ 观测zk和观测均值之差 ▷ innovation

4: $\mathbf{S}_k \leftarrow \mathbf{H}_k\boldsymbol{\Sigma}_k^-\mathbf{H}_k^\mathsf{T} + \mathbf{R}_k$ 观测zk的协方差 ▷ innovation covariance

5: $\mathbf{K}_k \leftarrow \boldsymbol{\Sigma}_k^-\mathbf{H}_k^\mathsf{T}\mathbf{S}_k^{-1}$ ▷ filter gain

6: $\boldsymbol{\mu}_k \leftarrow \boldsymbol{\mu}_k^- + \mathbf{K}_k\boldsymbol{\nu}_k$ ▷ corrected mean

7: $\boldsymbol{\Sigma}_k \leftarrow (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\boldsymbol{\Sigma}_k^-$ ▷ corrected covariance

8: $//\boldsymbol{\Sigma}_k \leftarrow (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\boldsymbol{\Sigma}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^\mathsf{T} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^\mathsf{T}$ ▷ numerically stable form

9: **return** $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$

根据条件高斯推得，见下图，
你会发现完全对应。

来自ECE564：theta=xk|k-1, x=zk

$$\underline{\mu}_{\theta|x} = \underline{\mu}_\theta + \boldsymbol{R}_\theta\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{R}_\theta\boldsymbol{H}^T + \boldsymbol{R})^{-1}(\underline{x} - \boldsymbol{H}\underline{\mu}_\theta)$$
$$= \underline{\mu}_\theta + (\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H} + \boldsymbol{R}_\theta^{-1})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\underline{x} - \boldsymbol{H}\underline{\mu}_\theta)$$
$$\boldsymbol{R}_{\theta|x} = \boldsymbol{R}_\theta - \boldsymbol{R}_\theta\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{R}_\theta\boldsymbol{H}^T + \boldsymbol{R})^{-1}\boldsymbol{H}\boldsymbol{R}_\theta$$
$$= (\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H} + \boldsymbol{R}_\theta^{-1})^{-1}$$

R_theta=Sigma_k^-

mu_k表示xk|xk-1,zk的均值
Sigma_k表示xk|xk-1,zk的协方差

Kalman Update: Innovation, Confidence, and Gain

Update step: 协方差推得的一切内容，比如新的协方差和卡尔曼增益和观测的具体值zk是无关的。

$$\mu_k \;=\; \mu_k^- \;+\; K_k\,\nu_k, \qquad \Sigma_k \;=\; (I - K_k H_k)\Sigma_k^-.$$

To remember:

$$\nu_k := z_k - H_k\mu_k^- \qquad \text{(innovation / residual)}$$

$$S_k := H_k\Sigma_k^- H_k^\top + R_k \qquad \text{(innovation covariance)}$$

$$K_k := \Sigma_k^- H_k^\top S_k^{-1} \qquad \text{(Kalman gain)}$$

- If the sensor is noisy (R_k large) $\Rightarrow$ S_k large $\Rightarrow$ K_k small $\Rightarrow$ *trust the model more.*
- If the prediction is uncertain (Σ_k^- large) $\Rightarrow$ K_k larger $\Rightarrow$ *trust the measurement more.*

Kalman gain chooses the optimal (MMSE) trade-off between prediction and measurement.

Probabilistic Derivation for the Multivariate Gaussian Case

In the prediction step, we use the joint distribution of the state at time steps k and $k-1$, $p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{u}_{1:k}, \mathbf{z}_{1:k-1})$, and marginalize $\mathbf{x}_{k-1}$ as follows:

$$\underbrace{\begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k-1} \end{bmatrix}}_{\mathbf{y}_k} = \underbrace{\begin{bmatrix} \mathbf{F}_k \\ \mathbf{I} \end{bmatrix}}_{\mathbf{A}_k} \mathbf{x}_{k-1} + \underbrace{\begin{bmatrix} \mathbf{G}_k \\ \mathbf{0} \end{bmatrix}}_{\mathbf{B}_k} \mathbf{u}_k + \underbrace{\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{W}_k} \mathbf{w}_k$$

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{W}_k \mathbf{w}_k$$

Probabilistic Derivation for the Multivariate Gaussian Case

We compute the mean and covariance as:

$$
\begin{aligned}
\mathbb{E}[\mathbf{y}_k] &= \mathbb{E}[\mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{W}_k \mathbf{w}_k] \\
&= \mathbf{A}_k \boldsymbol{\mu}_{k-1} + \mathbf{B}_k \mathbf{u}_k = \left[\begin{array}{c} \mathbf{F}_k \boldsymbol{\mu}_{k-1} + \mathbf{G}_k \mathbf{u}_k \\ \boldsymbol{\mu}_{k-1} \end{array} \right]
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}[\mathbf{y}_k] = \mathbb{E}[(\mathbf{y}_k - \mathbb{E}[\mathbf{y}_k])(\mathbf{y}_k - \mathbb{E}[\mathbf{y}_k])^\mathsf{T}] \\
&= \mathbb{E}[(\mathbf{A}_k(\mathbf{x}_{k-1} - \boldsymbol{\mu}_{k-1}) + \mathbf{W}_k \mathbf{w}_k)(\mathbf{A}_k(\mathbf{x}_{k-1} - \boldsymbol{\mu}_{k-1}) + \mathbf{W}_k \mathbf{w}_k)^\mathsf{T}] \\
&= \mathbf{A}_k \boldsymbol{\Sigma}_{k-1} \mathbf{A}_k^\mathsf{T} + \mathbf{W}_k \mathbf{Q}_k \mathbf{W}_k^\mathsf{T} \\
&= \left[\begin{array}{cc} \mathbf{F}_k \boldsymbol{\Sigma}_{k-1} \mathbf{F}_k^\mathsf{T} + \mathbf{Q}_k & \mathbf{F}_k \boldsymbol{\Sigma}_{k-1} \\ \boldsymbol{\Sigma}_{k-1} \mathbf{F}_k^\mathsf{T} & \boldsymbol{\Sigma}_{k-1} \end{array} \right]
\end{aligned}
$$

Probabilistic Derivation for the Multivariate Gaussian Case

Using the marginalization property of jointly Gaussian random vectors we have:

$$\mathbb{E}[\mathbf{x}_k] \triangleq \boldsymbol{\mu}_k^- = \mathbf{F}_k \boldsymbol{\mu}_{k-1} + \mathbf{G}_k \mathbf{u}_k$$

$$\mathrm{Cov}[\mathbf{x}_k] \triangleq \boldsymbol{\Sigma}_k^- = \mathbf{F}_k \boldsymbol{\Sigma}_{k-1} \mathbf{F}_k^\mathsf{T} + \mathbf{Q}_k$$

Probabilistic Derivation for the Multivariate Gaussian Case

In the correction step, we form the joint distribution $p(\mathbf{x}_k, \mathbf{z}_k | \mathbf{u}_{1:k}, \mathbf{z}_{1:t-1})$ and then condition on $\mathbf{z}_k$.

$$\underbrace{\begin{bmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix}}_{\mathbf{s}_k} = \underbrace{\begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix}}_{\mathbf{C}_k} \mathbf{x}_k + \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}}_{\mathbf{V}_k} \mathbf{v}_k$$

$$\mathbf{s}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{V}_k \mathbf{v}_k$$

Probabilistic Derivation for the Multivariate Gaussian Case

$$\mathbb{E}[\mathbf{s}_k] = \mathbb{E}[\mathbf{C}_k \mathbf{x}_k + \mathbf{V}_k \mathbf{v}_k] = \mathbf{C}_k \boldsymbol{\mu}_k^- = \left[\begin{array}{c} \boldsymbol{\mu}_k^- \\ \mathbf{H}_k \boldsymbol{\mu}_k^- \end{array} \right]$$

$$\begin{aligned} \mathrm{Cov}[\mathbf{s}_k] &= \mathbb{E}[(\mathbf{s}_k - \mathbb{E}[\mathbf{s}_k])(\mathbf{s}_k - \mathbb{E}[\mathbf{s}_k])^\mathsf{T}] \\ &= \mathbf{C}_k \boldsymbol{\Sigma}_k^- \mathbf{C}_k^\mathsf{T} + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^\mathsf{T} \\ &= \left[\begin{array}{cc} \boldsymbol{\Sigma}_k^- & \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T} \\ \mathbf{H}_k \boldsymbol{\Sigma}_k^- & \mathbf{H}_k \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T} + \mathbf{R}_k \end{array} \right] \end{aligned}$$

Probabilistic Derivation for the Multivariate Gaussian Case

Using the conditioning property of jointly Gaussian random vectors we have:

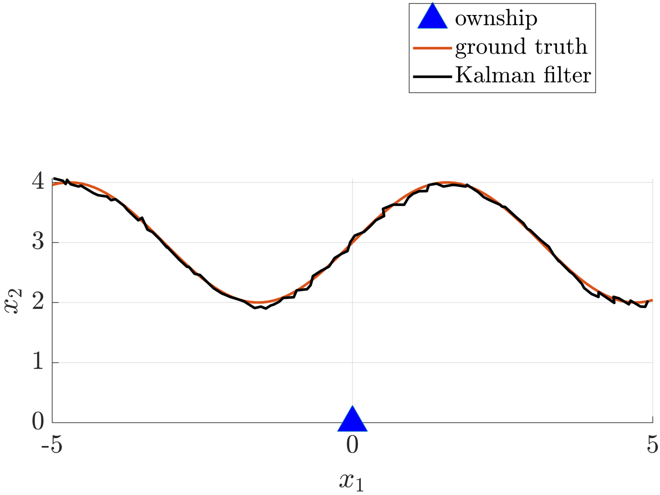
$$\mathbb{E}[\mathbf{x}_k|\mathbf{z}_k] \triangleq \boldsymbol{\mu}_k = \boldsymbol{\mu}_k^- + \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T}(\mathbf{H}_k \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T} + \mathbf{R}_k)^{-1}(\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k^-)$$

$$\begin{aligned} \mathrm{Cov}[\mathbf{x}_k|\mathbf{z}_k] \triangleq \boldsymbol{\Sigma}_k &= \boldsymbol{\Sigma}_k^- - \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T}(\mathbf{H}_k \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T} + \mathbf{R}_k)^{-1} \mathbf{H}_k \boldsymbol{\Sigma}_k^- \\ &= (\mathbf{I} - \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T}(\mathbf{H}_k \boldsymbol{\Sigma}_k^- \mathbf{H}_k^\mathsf{T} + \mathbf{R}_k)^{-1} \mathbf{H}_k) \boldsymbol{\Sigma}_k^- \end{aligned}$$

A target is moving in a 2D plane. The ownship position is known and fixed at the origin. We have access to noisy measurements that directly observe the target 2D coordinates at any time step.

$$\mathbf{F}_k = \mathbf{I}_2,\ \mathbf{G}_k = \mathbf{0}_2,\ \mathbf{H}_k = \mathbf{I}_2,\ \mathbf{Q}_k = 0.001\ \mathbf{I}_2,\ \mathbf{R}_k = 0.05^2\ \mathbf{I}_2$$

See `kf_single_target.m` for code.



- Under the Gaussian assumption for the initial state (or initial state error) and all the noises entering into the system, the Kalman filter is the optimal MMSE state estimator.

- If these random variables are not Gaussian and one has only their first two moments, then the Kalman filter algorithm is the best linear state estimator (Linear MMSE).

Minimum Mean Square Error (MMSE) Estimation

The MMSE estimation of $\mathbf{x}$ in terms of $\mathbf{z}$ is:

$$\hat{\mathbf{x}}^{\mathrm{MMSE}} = \arg\min_{\hat{\mathbf{x}}} \mathbb{E}[(\hat{\mathbf{x}} - \mathbf{x})^2 | \mathbf{z}]$$

The solution is the conditional mean:

$$\hat{\mathbf{x}}^{\mathrm{MMSE}} = \mathbb{E}[\mathbf{x}|\mathbf{z}] = \int \mathbf{x} p(\mathbf{x}|\mathbf{z}) \mathrm{d}\mathbf{x}$$

which can be obtained by

$$\frac{\partial \mathbb{E}[(\hat{\mathbf{x}} - \mathbf{x})^2 | \mathbf{z}]}{\partial \hat{\mathbf{x}}} = \mathbb{E}[2(\hat{\mathbf{x}} - \mathbf{x}) | \mathbf{z}] = 2(\hat{\mathbf{x}} - \mathbb{E}[\mathbf{x}|\mathbf{z}]) = 0$$

- $\hat{\mathbf{x}}$: Estimate
- $\mathbf{x}$: True value

Minimum Mean Square Error (MMSE) Estimation

MMSE estimation yields conditional mean:

$$\hat{\mathbf{x}}^{\mathrm{MMSE}} = \arg\min_{\hat{\mathbf{x}}} \mathbb{E}[(\hat{\mathbf{x}} - \mathbf{x})^2 | \mathbf{z}] = \mathbb{E}[\mathbf{x}|\mathbf{z}] = \int \mathbf{x} p(\mathbf{x}|\mathbf{z}) \mathrm{d}\mathbf{x}$$

Remark

Under the Gaussian assumption, the posterior mean estimated by the Kalman filter is exact, thus yielding the optimal MMSE estimation.

Remark

If the conditional PDF $p(\mathbf{x}|\mathbf{z})$ is Gaussian, then MMSE and Maximum a Posteriori estimators coincide since the mode and mean of the Gaussian distribution are the same.

Limitations of Kalman Filter

- Nonlinear motion (process) and measurement models;
- Unknown control inputs or mode changes;
- Data association uncertainty;
- Autocorrelated or crosscorrelated noise sequences.

Readings

- Probabilistic Robotics: Ch. 2 and Ch. 3
- State Estimation for Robotics: Ch. 3 and Ch. 4
- Lecture notes 2 and 3