Kalman Filter

Math 628

November 14, 2025

Outline

State-Space Model

We work in discrete time $k = 0, 1, 2, \dots$

State dynamics:

$$x_{k+1} = Ax_k + Bu_k + w_k$$

Measurement model:

$$y_k = Cx_k + v_k$$

Assumptions:

- $x_k \in \mathbb{R}^n$: state (hidden)
- $u_k \in \mathbb{R}^m$: known input (treated as deterministic)
- $y_k \in \mathbb{R}^p$: measurement
- $w_k \sim \mathcal{N}(0, Q), \quad v_k \sim \mathcal{N}(0, R)$
- $\{w_k\}, \{v_k\}$ independent over time and mutually independent
- Initial state:

$$x_0 \sim \mathcal{N}(\hat{x}_{0|0}, P_{0|0})$$

Filtering Goal

Define the measurement history:

$$\mathcal{Y}_k := \{y_1, y_2, \dots, y_k\}.$$

Objective:

- Compute the posterior distribution

$$p(x_k \mid \mathcal{Y}_k).$$

- Under linear–Gaussian assumptions:

$$x_k \mid \mathcal{Y}_k \sim \mathcal{N}(\hat{x}_{k|k}, P_{k|k}).$$

- The MMSE estimate:

$$\hat{x}_{k|k} := \mathbb{E}[x_k \mid \mathcal{Y}_k].$$

Kalman filter: Recursive computation of $(\hat{x}_{k|k}, P_{k|k})$ as new measurements arrive.

Gaussian Random Vectors

Let $z \in \mathbb{R}^n$ be a random vector.

Definition: z is Gaussian with mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{R}^{n \times n}$, written

$$z \sim \mathcal{N}(\mu, \Sigma),$$

if its pdf is

$$p(z) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\tfrac{1}{2}(z-\mu)^\top \Sigma^{-1}(z-\mu)\right).$$

We will use several key properties:

- Linear transformation of Gaussians.
- Joint distribution under linear relations.
- Conditional distribution of a joint Gaussian.

Property 1: Linear Transformation of a Gaussian

Theorem (Linear transformation).
If $z \sim \mathcal{N}(\mu, \Sigma)$ and

$$x = Fz + g,$$

with matrix F and constant vector g, then

$$x \sim \mathcal{N}\big(F\mu + g,\ F\Sigma F^\top\big).$$

Proof.

- Mean:

$$\mathbb{E}[x] = \mathbb{E}[Fz + g] = F\mathbb{E}[z] + g = F\mu + g.$$

- Covariance:

$$\mathrm{Cov}(x) = \mathbb{E}\left[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^\top\right].$$

  Substitute $x = Fz + g$, $\mathbb{E}[x] = F\mu + g$:

$$x - \mathbb{E}[x] = F(z - \mu),$$

  so

$$\mathrm{Cov}(x) = F\mathbb{E}\left[(z - \mu)(z - \mu)^\top\right] F^\top = F\Sigma F^\top.$$

Property 2: Joint Gaussian from Linear Relation

Let

$$x \sim \mathcal{N}(m, P), \qquad v \sim \mathcal{N}(0, R),$$

with x and v independent.

Define a linear measurement:

$$y = Cx + v.$$

Claim:

$$\begin{bmatrix} x \\ y \end{bmatrix} \text{ is jointly Gaussian.}$$

Reason:

- The stacked vector $\begin{bmatrix} x \\ v \end{bmatrix}$ is Gaussian (independent Gaussians).
- We can write

$$\begin{bmatrix} x \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} I \\ C \end{bmatrix}}_{F_1} x + \underbrace{\begin{bmatrix} 0 \\ I \end{bmatrix}}_{F_2} v,$$

  an affine (linear + constant) function of Gaussian variables.
- By the linear transformation property, the result is Gaussian.

Joint Mean and Covariance of (x, y)

We now compute mean and covariance.

Means:

$$\mathbb{E}[x] = m, \qquad \mathbb{E}[y] = \mathbb{E}[Cx + v] = Cm + 0 = Cm.$$

Thus

$$\mathbb{E}\left[\begin{bmatrix} x \\ y \end{bmatrix}\right] = \begin{bmatrix} m \\ Cm \end{bmatrix}.$$

Covariances:

$$\mathrm{Cov}(x, x) = P.$$

$$\mathrm{Cov}(x, y) = \mathbb{E}[(x - m)(y - Cm)^\top] = \mathbb{E}[(x - m)(C(x - m) + v)^\top].$$

Use independence: $\mathbb{E}[(x - m)v^\top] = 0$:

$$\mathrm{Cov}(x, y) = \mathbb{E}[(x - m)(x - m)^\top C^\top] = PC^\top.$$

Similarly:

$$\mathrm{Cov}(y, x) = CP, \qquad \mathrm{Cov}(y, y) = CPC^\top + R.$$

So

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m \\ Cm \end{bmatrix}, \begin{bmatrix} P & PC^\top \\ CP & CPC^\top + R \end{bmatrix}\right).$$

Conditional Gaussian: Statement

Theorem (Conditional of a joint Gaussian).
Suppose

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_x \\ m_y \end{bmatrix}, \begin{bmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{bmatrix} \right),$$

with P_{yy} invertible.

Then the conditional distribution of x given y is Gaussian:

$$x \mid y \sim \mathcal{N}(\tilde{m}_x, \tilde{P}_{xx}),$$

where

$$\tilde{m}_x = m_x + P_{xy} P_{yy}^{-1} (y - m_y),$$

$$\tilde{P}_{xx} = P_{xx} - P_{xy} P_{yy}^{-1} P_{yx}.$$

This result underlies the *measurement update* in the Kalman filter.

Conditional Gaussian: Proof Idea

We start from the joint pdf:

$$p(x, y) = \frac{1}{\sqrt{(2\pi)^{n+m} \det \Sigma}} \exp\left(-\frac{1}{2} \begin{bmatrix} x - m_x \\ y - m_y \end{bmatrix}^\top \Sigma^{-1} \begin{bmatrix} x - m_x \\ y - m_y \end{bmatrix} \right),$$

with

$$\Sigma = \begin{bmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{bmatrix}.$$

We want

$$p(x \mid y) = \frac{p(x, y)}{p(y)}.$$

For a fixed y, $p(y)$ is a normalization constant, so $p(x \mid y)$ is proportional to the exponential term, viewed as a function of x.

We will:

- Find the exponent as a quadratic function in x.
- "Complete the square" to identify a Gaussian in x.

Block Inverse (Schur Complement)

We use a standard block inverse formula. Let

$$\Sigma = \begin{bmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{bmatrix},$$

with P_{yy} invertible.

Define the Schur complement:

$$S := P_{xx} - P_{xy} P_{yy}^{-1} P_{yx}.$$

Then

$$\Sigma^{-1} = \begin{bmatrix} S^{-1} & -S^{-1} P_{xy} P_{yy}^{-1} \\ -P_{yy}^{-1} P_{yx} S^{-1} & P_{yy}^{-1} + P_{yy}^{-1} P_{yx} S^{-1} P_{xy} P_{yy}^{-1} \end{bmatrix}.$$

(Proof: verify $\Sigma^{-1}\Sigma = I$ by direct block multiplication.)

Conditional Gaussian: Completing the Square

Let

$$z = \begin{bmatrix} x - m_x \\ y - m_y \end{bmatrix}.$$

Then

$$-\tfrac{1}{2} z^\top \Sigma^{-1} z$$

is the exponent of the joint pdf.

Write Σ^{-1} in block form:

$$\Sigma^{-1} = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix},$$

where

$$A = S^{-1}, \quad B = -S^{-1} P_{xy} P_{yy}^{-1}, \quad D = P_{yy}^{-1} + P_{yy}^{-1} P_{yx} S^{-1} P_{xy} P_{yy}^{-1}.$$

Then

$$z^\top \Sigma^{-1} z = (x - m_x)^\top A (x - m_x) + 2(x - m_x)^\top B (y - m_y) + (y - m_y)^\top D (y - m_y).$$

For fixed y, terms involving only y go into $p(y)$; terms involving x define $p(x \mid y)$.

Conditional Gaussian: Mean and Covariance

Consider the quadratic form in x:

$$(x - m_x)^\top A(x - m_x) + 2(x - m_x)^\top B(y - m_y) + \text{(terms independent of } x\text{)}.$$

Complete the square:

$$\begin{aligned}&(x - m_x)^\top A(x - m_x) + 2(x - m_x)^\top B(y - m_y)\\ &= (x - \tilde{m}_x)^\top A(x - \tilde{m}_x) + \text{(terms independent of } x\text{)},\end{aligned}$$

for a suitable $\tilde{m}_x$.

Solving for $\tilde{m}_x$ gives

$$\tilde{m}_x = m_x - A^{-1}B(y - m_y).$$

Recall $A = S^{-1}$, $B = -S^{-1}P_{xy}P_{yy}^{-1}$, so

$$\tilde{m}_x = m_x + SP_{xy}P_{yy}^{-1}(y - m_y).$$

Using $S = P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}$, one can show

$$SP_{xy}P_{yy}^{-1} = P_{xy}P_{yy}^{-1},$$

which yields

$$\tilde{m}_x = m_x + P_{xy}P_{yy}^{-1}(y - m_y).$$

Bayesian Filtering Recap

Our goal is to recursively compute $p(x_k \mid \mathcal{Y}_k)$.

Two fundamental steps:

1. Prediction (time update):

$$p(x_{k+1} \mid \mathcal{Y}_k) = \int p(x_{k+1} \mid x_k)\, p(x_k \mid \mathcal{Y}_k)\, dx_k.$$

2. Update (measurement update):

$$p(x_{k+1} \mid \mathcal{Y}_{k+1}) \propto p(y_{k+1} \mid x_{k+1})\, p(x_{k+1} \mid \mathcal{Y}_k).$$

In the linear–Gaussian case, both densities remain Gaussian $\Rightarrow$ closed-form recursion for mean and covariance.

Induction Hypothesis

We proceed by induction on k.

Hypothesis:

$$x_k \mid \mathcal{Y}_k \sim \mathcal{N}(\hat{x}_{k|k}, P_{k|k}).$$

Base case: Given

$$x_0 \sim \mathcal{N}(\hat{x}_{0|0}, P_{0|0}),$$

the hypothesis holds for $k = 0$.

We show:

- Prediction:

$$x_{k+1} \mid \mathcal{Y}_k \sim \mathcal{N}(\hat{x}_{k+1|k}, P_{k+1|k}).$$

- Update:

$$x_{k+1} \mid \mathcal{Y}_{k+1} \sim \mathcal{N}(\hat{x}_{k+1|k+1}, P_{k+1|k+1}).$$

This proves Gaussianity for all k and gives the Kalman filter recursion.

Prediction Step: Setup

State equation:

$$x_{k+1} = Ax_k + Bu_k + w_k.$$

Given $\mathcal{Y}_k$:

- $x_k \mid \mathcal{Y}_k \sim \mathcal{N}(\hat{x}_{k|k}, P_{k|k})$.
- $w_k \sim \mathcal{N}(0, Q)$, independent of x_k and $\mathcal{Y}_k$.
- u_k is known (deterministic).

Then

$$x_{k+1} \mid \mathcal{Y}_k = Ax_k + Bu_k + w_k.$$

We want the distribution of x_{k+1} conditioned on $\mathcal{Y}_k$.

Prediction Step: Mean and Covariance

Mean:

$$\hat{x}_{k+1|k} := \mathbb{E}[x_{k+1} \mid \mathcal{Y}_k] = \mathbb{E}[Ax_k + Bu_k + w_k \mid \mathcal{Y}_k]$$
$$= A\mathbb{E}[x_k \mid \mathcal{Y}_k] + Bu_k + \mathbb{E}[w_k] = A\hat{x}_{k|k} + Bu_k.$$

Covariance:

$$\begin{aligned} P_{k+1|k} &:= \mathsf{Cov}(x_{k+1} \mid \mathcal{Y}_k) \\ &= \mathsf{Cov}(Ax_k + w_k \mid \mathcal{Y}_k) \\ &= A\mathsf{Cov}(x_k \mid \mathcal{Y}_k)A^\top + \mathsf{Cov}(w_k) \\ &= AP_{k|k}A^\top + Q. \end{aligned}$$

Since x_{k+1} is an affine transformation of Gaussian variables, we have

$$x_{k+1} \mid \mathcal{Y}_k \sim \mathcal{N}(\hat{x}_{k+1|k}, P_{k+1|k}).$$

This completes the prediction step. □

Update Step: Setup

Measurement equation at time $k+1$:

$$y_{k+1} = Cx_{k+1} + v_{k+1}, \quad v_{k+1} \sim \mathcal{N}(0, R).$$

From the prediction step:

$$x_{k+1} \mid \mathcal{Y}_k \sim \mathcal{N}(\hat{x}_{k+1|k}, P_{k+1|k}).$$

Noise v_{k+1} is independent of x_{k+1} and $\mathcal{Y}_k$.

Goal:

$$p(x_{k+1} \mid \mathcal{Y}_{k+1}) = p(x_{k+1} \mid \mathcal{Y}_k, y_{k+1}).$$

We will use the joint Gaussian of (x_{k+1}, y_{k+1}) given $\mathcal{Y}_k$ and apply the conditional Gaussian theorem.

Joint Distribution of (x_{k+1}, y_{k+1})

Given $\mathcal{Y}_k$:

- $x_{k+1} \sim \mathcal{N}(\hat{x}_{k+1|k}, P_{k+1|k})$.
- $y_{k+1} = Cx_{k+1} + v_{k+1}$, with $v_{k+1} \sim \mathcal{N}(0, R)$ independent.

Therefore

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} \Bigg| \mathcal{Y}_k \sim \mathcal{N}\left(\begin{bmatrix} \hat{x}_{k+1|k} \\ C\hat{x}_{k+1|k} \end{bmatrix}, \begin{bmatrix} P_{k+1|k} & P_{k+1|k}C^\top \\ CP_{k+1|k} & CP_{k+1|k}C^\top + R \end{bmatrix} \right).$$

Apply the conditional Gaussian theorem with:

$$m_x = \hat{x}_{k+1|k}, \quad m_y = C\hat{x}_{k+1|k},$$

$$P_{xx} = P_{k+1|k}, \quad P_{xy} = P_{k+1|k}C^\top, \quad P_{yy} = CP_{k+1|k}C^\top + R.$$

Update Step: Conditional Mean

By the conditional Gaussian formula:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + P_{k+1|k}C^\top \big(CP_{k+1|k}C^\top + R\big)^{-1}\big(y_{k+1} - C\hat{x}_{k+1|k}\big).$$

Define the **Kalman gain**:

$$K_{k+1} := P_{k+1|k}C^\top \big(CP_{k+1|k}C^\top + R\big)^{-1}.$$

Then:

$$\boxed{\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\big(y_{k+1} - C\hat{x}_{k+1|k}\big)}$$

The term

$$\tilde{y}_{k+1} := y_{k+1} - C\hat{x}_{k+1|k}$$

is called the *innovation* (measurement minus prediction).

Update Step: Conditional Covariance

Again from the conditional Gaussian formula:

$$P_{k+1|k+1} = P_{k+1|k} - P_{k+1|k}C^\top \left(CP_{k+1|k}C^\top + R\right)^{-1} CP_{k+1|k}.$$

Using the definition of the Kalman gain K_{k+1}:

$$K_{k+1} = P_{k+1|k}C^\top \left(CP_{k+1|k}C^\top + R\right)^{-1},$$

we obtain

$$\boxed{P_{k+1|k+1} = P_{k+1|k} - K_{k+1}CP_{k+1|k}}$$

or equivalently

$$P_{k+1|k+1} = (I - K_{k+1}C)P_{k+1|k}.$$

(There is also a symmetric "Joseph" form, but this is the standard compact one.)

Kalman Gain as Optimal Linear Gain (Sketch)

Consider a general linear update of the form

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K\big(y_{k+1} - C\hat{x}_{k+1|k}\big),$$

with some matrix K.

Define the estimation error:

$$e_{k+1|k+1} := x_{k+1} - \hat{x}_{k+1|k+1}.$$

One can show:

$$P_{k+1|k+1}(K) := \mathsf{Cov}(e_{k+1|k+1}) = (I - KC)P_{k+1|k}(I - KC)^\top + KRK^\top.$$

Minimizing $\operatorname{tr}\big(P_{k+1|k+1}(K)\big)$ w.r.t. K:

$$\frac{\partial}{\partial K}\operatorname{tr}\big(P_{k+1|k+1}(K)\big) = 0 \quad \Rightarrow \quad K = P_{k+1|k}C^\top(CP_{k+1|k}C^\top + R)^{-1}.$$

Thus the Kalman gain is the *optimal* linear gain (in the MMSE sense).

Kalman Filter: Full Recursion

Initialization:

$$\hat{x}_{0|0}, \quad P_{0|0}$$

given (or chosen).

For $k = 0, 1, 2, \ldots$ perform:

Prediction (Time Update):

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k,$$

$$P_{k+1|k} = AP_{k|k}A^\top + Q.$$

Update (Measurement Update):

$$\tilde{y}_{k+1} = y_{k+1} - C\hat{x}_{k+1|k},$$

$$S_{k+1} = CP_{k+1|k}C^\top + R,$$

$$K_{k+1} = P_{k+1|k}C^\top S_{k+1}^{-1},$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\tilde{y}_{k+1},$$

$$P_{k+1|k+1} = (I - K_{k+1}C)P_{k+1|k}.$$

Summary

- Linear–Gaussian state-space model:

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad y_k = Cx_k + v_k.$$

- Assumption: all relevant variables are Gaussian.
- By induction, $x_k \mid \mathcal{Y}_k$ remains Gaussian.
- **Prediction step**: use system dynamics to propagate mean and covariance.
- **Update step**: use the conditional Gaussian formula to incorporate new measurements.
- The Kalman filter is the optimal (MMSE) linear recursive estimator under these assumptions.

Questions?