Kalman Filter in Three Ways

Geometric, Probabilistic, and Optimization Perspectives

Zirui Zhang

Cheng Kar-Shun Robotics Institute
The Hong Kong University of Science and Technology

September 21, 2025

System Model and Assumptions

Consider a discrete-time linear Gaussian system with initial condition x_0 and P_0:

$$
\begin{aligned}
x_{k+1} &= A_k x_k + B_k u_k + \omega_k, \quad & \omega_k &\sim \mathcal{N}(0, Q_k) \\
y_k &= C_k x_k + \nu_k, & \nu_k &\sim \mathcal{N}(0, R_k)
\end{aligned}
$$

Assumptions:

- (A_k, B_k) is *controllable* and (A_k, C_k) is *observable*
- $Q_k \succeq 0, R_k \succeq 0, P_0 \succeq 0$
- ω_k, ν_k and x_0 are mutually uncorrelated
- The future state of the system is conditionally independent of the past states given the current state

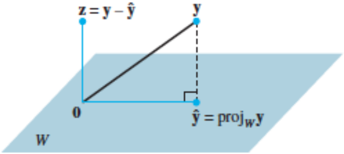
Goal: Find $\hat{x}_{k|k} = \mathbb{E}[x_k | y_{1:k}]$ (MMSE estimator)

Hilbert Space of Random Variables

Key Idea:

- View random variables as vectors in Hilbert space
- Inner product: $\langle \xi, \eta \rangle = \mathbb{E}[\xi\eta]$
- Orthogonality: $\xi \perp \eta \Leftrightarrow \mathbb{E}[\xi\eta] = 0$
- Optimal estimate is orthogonal projection onto observation space

Geometric Interpretation:



Time Update

State Prediction:

$$
\begin{aligned}
\hat{x}_{k|k-1} &= \mathbb{E}[x_k \mid y_{1:k-1}] \\
&= \mathbb{E}[A_{k-1} x_{k-1} + B_{k-1} u_{k-1} + w_{k-1} \mid y_{1:k-1}] \\
&= A_{k-1} \hat{x}_{k-1|k-1} + B_{k-1} u_{k-1} \qquad (\text{since } w_{k-1} \perp y_{1:k-1})
\end{aligned}
$$

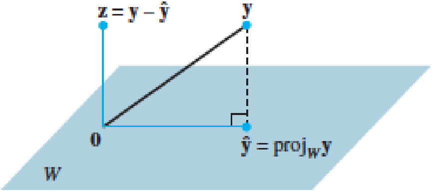
Covariance Prediction:

$$
\begin{aligned}
P_{k|k-1} &= \operatorname{cov}(x_k - \hat{x}_{k|k-1}) \\
&= \operatorname{cov}[A_{k-1}(x_{k-1} - \hat{x}_{k-1|k-1}) + w_{k-1}] \\
&= A_{k-1} \cdot \operatorname{cov}(x_{k-1} - \hat{x}_{k-1|k-1}) \cdot A_{k-1}^\top + 2A_{k-1} \cdot \operatorname{cov}(x_k - \hat{x}_{k|k-1}, \omega_{k-1}) + \operatorname{cov}(w_{k-1}) \\
&= A_{k-1} P_{k-1|k-1} A_{k-1}^\top + Q_{k-1}
\end{aligned}
$$

Innovation Process

Definition:

$$
\begin{aligned}
e_k &= y_k - \hat{y}_{k|k-1} \\
&= y_k - \mathrm{proj}_{\mathcal{Y}_{k-1}}(y_k) \\
&= y_k - \mathrm{proj}_{\mathcal{Y}_{k-1}}(C_k x_k + v_k) \\
&= y_k - C_k \cdot \mathrm{proj}_{\mathcal{Y}_{k-1}}(x_k) - \mathrm{proj}_{\mathcal{Y}_{k-1}}(v_k) \\
&= y_k - C_k \hat{x}_{k|k-1}
\end{aligned}
$$



Properties:

- Zero Mean: $\mathbb{E}[e_k] = 0$
- White Sequence: $\mathbb{E}[e_k e_j^\top] = 0$ for $k \neq j$
- Orthogonality Principle: $\mathbb{E}[e_k y_j^\top] = 0$ for $j < k$

Measurement Update

State Update:

$$\begin{aligned}\hat{x}_{k|k} &= \mathrm{proj}_{\mathcal{Y}_k}(x_k) \\ &= \hat{x}_{k|k-1} + K_k e_k \\ &= \hat{x}_{k|k-1} + K_k(y_k - C_k \hat{x}_{k|k-1})\end{aligned}$$

Covariance Update:

$$\begin{aligned}P_{k|k} &= \mathrm{cov}(x_k - \hat{x}_{k|k}) \\ &= \mathrm{cov}(x_k - \hat{x}_{k|k-1} - K_k e_k) \\ &= \mathrm{cov}(x_k - \hat{x}_{k|k-1}) - 2K_k \mathrm{cov}(x_k - \hat{x}_{k|k-1}, e_k) + K_k \mathrm{cov}(e_k) K_k^\top \\ &= \mathrm{cov}(x_k - \hat{x}_{k|k-1}) - 2K_k \mathrm{cov}(x_k - \hat{x}_{k|k-1}, y_k - C_k \hat{x}_{k|k-1}) + K_k \mathrm{cov}(y_k - C_k \hat{x}_{k|k-1}) K_k^\top \\ &= P_{k|k-1} - K_k C_k P_{k|k-1} - P_{k|k-1} C_k^\top K^\top + K_k (C_k P_{k|k-1} C_k^\top + R_k) K_k^\top\end{aligned}$$

Kalman Gain Derivation

Optimal Kalman Gain:

$$\frac{\partial \mathrm{tr}(P_{k|k})}{\partial K_k} = -2P_{k|k-1}C_k^\top + 2K_k(C_kP_{k|k-1}C_k^\top + R_k) = 0$$

$$K_k = P_{k|k-1}C_k^\top(C_kP_{k|k-1}C_k^\top + R_k)^{-1}$$

Covariance Derivation:

$$P_{k|k} = P_{k|k-1} - K_kC_kP_{k|k-1} = (P_{k|k-1}^{-1} + C_k^\top R_k^{-1}C_k)^{-1}$$

Bayesian Filtering Framework

$$
\begin{aligned}
& p(x_k|y_{1:k}, u_{1:k}) \\
&= p(x_k|y_k, y_{1:k-1}, u_{1:k}) \\
&= \frac{p(y_k|x_k, y_{1:k-1}, u_{1:k}) \cdot p(x_k|y_{1:k-1}, u_{1:k})}{p(y_k|y_{1:k-1}, u_{1:k})} \\
&= \eta \cdot p(y_k|x_k) \cdot p(x_k|y_{1:k-1}, u_{1:k}) \\
&= \eta \cdot p(y_k|x_k) \cdot \int p(x_k, x_{k-1}|y_{1:k-1}, u_{1:k}) \,\mathrm{d}x_{k-1} \\
&= \eta \cdot p(y_k|x_k) \cdot \int p(x_k|x_{k-1}, y_{1:k-1}, u_{1:k}) \cdot p(x_{k-1}|y_{1:k-1}, u_{1:k}) \,\mathrm{d}x_{k-1} \\
&= \eta \cdot \underbrace{p(y_k|x_k)}_{\text{observation model}} \cdot \int \underbrace{p(x_k|x_{k-1}, u_k)}_{\text{motion model}} \cdot \underbrace{p(x_{k-1}|y_{1:k-1}, u_{1:k-1})}_{\text{previous belief}} \,\mathrm{d}x_{k-1}
\end{aligned}
$$

Prediction Step: Gaussian Propagation

$$p(x_k|y_{1:k}, u_{1:k}) = \eta \cdot \mathcal{N}(y_k; C_k x_k, R_k) \cdot \int \mathcal{N}(x_k; A_{k-1}x_{k-1} + B_{k-1}u_{k-1}, Q_{k-1}) \cdot \mathcal{N}(x_{k-1}; \hat{x}_{k-1}, P_{k-1})\, \mathrm{d}x_{k-1}$$

Predicted Mean:

$$\begin{aligned}\hat{x}_{k|k-1} &= \mathbb{E}[A_{k-1}x_{k-1} + B_{k-1}u_{k-1} + w_{k-1}] \\ &= A_{k-1}\mathbb{E}[x_{k-1}] + B_{k-1}u_{k-1} + \mathbb{E}[w_{k-1}] \\ &= A_{k-1}\hat{x}_{k-1} + B_{k-1}u_{k-1}\end{aligned}$$

Predicted Covariance:

$$\begin{aligned}P_{k|k-1} &= \mathrm{cov}[A_{k-1}x_{k-1} + B_{k-1}u_{k-1} + w_{k-1}] \\ &= \mathrm{cov}[A_{k-1}x_{k-1}] + \mathrm{cov}[w_{k-1}] \\ &= A_{k-1}\mathrm{cov}[x_{k-1}]A_{k-1}^\top + Q_{k-1} \\ &= A_{k-1}P_{k-1}A_{k-1}^\top + Q_{k-1}\end{aligned}$$

Update Step: Gaussian Product

$$p(x_k|y_{1:k}, u_{1:k}) = \eta \cdot \mathcal{N}(y_k; C_k x_k, R_k) \cdot \mathcal{N}(x_k; \hat{x}_{k|k-1}, P_{k|k-1})$$

Gaussian Product:

$$\mathcal{N}(x; \mu, \Sigma) \propto \mathcal{N}(x; \mu_1, \Sigma_1) \cdot \mathcal{N}(x, \mu_2, \Sigma_2)$$

$$\begin{aligned} \Sigma^{-1} &= \Sigma_1^{-1} + \Sigma_2^{-1} \\ \mu &= \Sigma(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2) \end{aligned}$$

Posterior Result:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - C_k \hat{x}_{k|k-1}) \\ K_k &= P_{k|k-1} C_k^\top (C_k P_{k|k-1} C_k^\top + R)^{-1} \\ P_{k|k} &= (I - K_k C_k) P_{k|k-1} \end{aligned}$$

Maximum A Posteriori Formulation

MAP Estimation:

$$\begin{aligned}\hat{x}_{k|k} &= \arg\max_{x_k} p(x_k \mid y_{1:k}) \\ &= \arg\min_{x_k} \left[-\log p(x_k \mid y_{1:k})\right]\end{aligned}$$

Weighted Least Square:

$$\mathcal{E}(x) = ||A_{k-1}x - b||_{\Sigma}^{2} = x^{\top}A_{k-1}^{\top}\Sigma^{-1}A_{k-1}x - 2b^{\top}\Sigma^{-1}A_{k-1}x + b^{\top}\Sigma^{-1}b$$

$$\nabla\mathcal{E} = 2A_{k-1}^{\top}\Sigma^{-1}A_{k-1}x - 2A_{k-1}^{\top}\Sigma^{-1}b$$

$$\hat{x} = (A_{k-1}^{\top}\Sigma^{-1}A_{k-1})^{-1}A_{k-1}^{\top}\Sigma^{-1}b$$

MAP as Weighted Least Squares

Posterior Distribution:

$$p(x_k \mid y_{1:k}) \propto p(y_k \mid x_k)\, p(x_k \mid y_{1:k-1})$$

Assume Gaussian Distributions:

$$p(x_k \mid y_{1:k-1}) = \mathcal{N}(x_k; \hat{x}_{k|k-1}, P_{k|k-1})$$
$$p(y_k \mid x_k) = \mathcal{N}(y_k; C_k x_k, R_k)$$

Negative Log-Posterior:

$$-\log p(x_k \mid y_{1:k}) \propto \frac{1}{2}\|y_k - C_k x_k\|^2_{R_k^{-1}} + \frac{1}{2}\|x_k - \hat{x}_{k|k-1}\|^2_{P_{k|k-1}^{-1}}$$
$$= \frac{1}{2}\left\| \begin{bmatrix} C_k \\ I \end{bmatrix} x_k - \begin{bmatrix} y_k \\ \hat{x}_{k|k-1} \end{bmatrix} \right\|^2_{\Sigma^{-1}}$$

where $\Sigma = \begin{bmatrix} R_k & 0 \\ 0 & P_{k|k-1} \end{bmatrix}$.

MAP Solution

Weighted Least Squares Form:

$$A_{k-1} = \begin{bmatrix} C_k \\ I \end{bmatrix}, \quad b = \begin{bmatrix} y_k \\ \hat{x}_{k|k-1} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} R_k & 0 \\ 0 & P_{k|k-1} \end{bmatrix}$$

MAP Estimate:

$$\begin{aligned} \hat{x}_{k|k} &= \left(A_{k-1}^\top \Sigma^{-1} A_{k-1}\right)^{-1} A_{k-1}^\top \Sigma^{-1} b \\ &= \left(C_k^\top R_k^{-1} C_k + P_{k|k-1}^{-1}\right)^{-1} \left(C_k^\top R_k^{-1} y_k + P_{k|k-1}^{-1} \hat{x}_{k|k-1}\right) \end{aligned}$$

Equivalence Proof

Using Matrix Inversion Lemma:

$$\begin{aligned}\hat{x}_{k|k} &= \left(C_k^\top R_k^{-1} C_k + P_{k|k-1}^{-1}\right)^{-1} \left(C_k^\top R_k^{-1} y_k + P_{k|k-1}^{-1} \hat{x}_{k|k-1}\right) \\ &= \hat{x}_{k|k-1} + P_{k|k-1} C_k^\top (C_k P_{k|k-1} C_k^\top + R_k)^{-1} (y_k - C_k \hat{x}_{k|k-1})\end{aligned}$$

Proof:

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

This shows the equivalence between the MAP solution and the Kalman update.

Theoretical Insights and Extensions

Key Insights:

- **Geometric**: Reveals orthogonality principle and innovation process
- **Probabilistic**: Shows optimality under Gaussian assumptions
- **Optimization**: Connects to weighted least squares and regularization

Unified Algorithm: All approaches yield the same recursive equations:

$$\text{time update}\begin{cases}\hat{x}_{k|k-1} = A_{k-1}\hat{x}_{k-1|k-1}\\ P_{k|k-1} = A_{k-1}P_{k-1|k-1}A_{k-1}^\top + Q\end{cases}$$

$$\text{measurement update}\begin{cases}K_k = P_{k|k-1}C_k^\top(C_kP_{k|k-1}C_k^\top + R)^{-1}\\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - C_k\hat{x}_{k|k-1})\\ P_{k|k} = (I - K_kC_k)P_{k|k-1}\end{cases}$$

Extensions:

- Nonlinear systems: EKF, UKF, particle filters
- Non-Gaussian noise: robust Kalman filters

Thank you for listening !

Zirui Zhang