Extended Kalman Filter - Derivations & Proofs

Paul F. Roysdon, Ph.D.

Contents

1 Mathematical Derivations & Proofs

1.1 Introduction

The (discrete-time) Kalman Filter (KF) estimates the latent state of a linear dynamical system from noisy measurements by recursively applying Bayes' rule under Gaussian assumptions. Because linear transforms preserve Gaussianity and the product of Gaussians is Gaussian, the exact posterior mean and covariance admit closed forms. The Extended Kalman Filter (EKF) generalizes KF to smooth nonlinear models by locally linearizing the dynamics and the observation map.

We derive (i) the linear-Gaussian state-space model; (ii) the KF *prediction* and *update* equations for the mean and covariance; (iii) the *Kalman gain* via two routes: product-of-Gaussians and minimum mean-square error; (iv) numerically stable covariance updates (Joseph form); (v) the EKF via first-order Taylor expansions and Jacobians; and (vi) complete algorithms with dimensions and assumptions explicitly stated.

1.2 Data and Notation

Consider time indices $t = 1, 2, \ldots, T$. Let

$$\mathbf{x}_t \in \mathbb{R}^n \quad \text{(state)}, \qquad \mathbf{y}_t \in \mathbb{R}^p \quad \text{(measurement)}, \qquad \mathbf{u}_t \in \mathbb{R}^m \quad \text{(control, optional)}.$$

Matrices (possibly time-varying) and noise statistics:

$$\mathbf{A}_t \in \mathbb{R}^{n\times n},\ \mathbf{B}_t \in \mathbb{R}^{n\times m},\ \mathbf{C}_t \in \mathbb{R}^{p\times n}, \quad \mathbf{Q}_t \succeq \mathbf{0} \in \mathbb{R}^{n\times n},\ \mathbf{R}_t \succeq \mathbf{0} \in \mathbb{R}^{p\times p}.$$

Let $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a Gaussian with mean $\boldsymbol{\mu}$, covariance $\boldsymbol{\Sigma}$. We use $\widehat{\mathbf{x}}_{t|s}$ and $\mathbf{P}_{t|s}$ for the posterior mean and covariance at time t given measurements up to time s ($s \in \{t-1, t\}$). $\mathbf{I}_n$ is the $n \times n$ identity.

1.3 Model Formulation (Linear-Gaussian State Space)

The *process* and *measurement* models are

$$\text{State evolution:} \quad \mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{B}_t\mathbf{u}_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \tag{1}$$

$$\text{Measurement:} \quad \mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t). \tag{2}$$

Assumptions. $(\mathbf{x}_0, \{\mathbf{w}_t\}, \{\mathbf{v}_t\})$ are jointly Gaussian, mutually independent across time and between process/measurement; $\mathbf{x}_0 \sim \mathcal{N}(\widehat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$ is given.

Under Eqns. (1)–(2), the one-step predictive prior and the likelihood are Gaussian; hence the posterior $p(\mathbf{x}_t \mid \mathbf{y}_{1:t})$ is Gaussian and fully characterized by its mean/covariance.

1.4 Kalman Filter: Bayesian Derivation (Product of Gaussians)

Assume at time $t-1$ we have the posterior $\mathbf{x}_{t-1} \,|\, \mathbf{y}_{1:t-1} \sim \mathcal{N}(\widehat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1})$.

Prediction (Time Update). By linearity and independence,

$$\widehat{\mathbf{x}}_{t|t-1} = \mathbf{A}_t\widehat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}_t\mathbf{u}_t, \tag{3}$$

$$\mathbf{P}_{t|t-1} = \mathbf{A}_t\mathbf{P}_{t-1|t-1}\mathbf{A}_t^\top + \mathbf{Q}_t. \tag{4}$$

The predictive likelihood of $\mathbf{y}_t$ is $\mathcal{N}(\mathbf{C}_t\widehat{\mathbf{x}}_{t|t-1},\ \mathbf{S}_t)$ with

$$\mathbf{S}_t \;=\; \mathbf{C}_t\mathbf{P}_{t|t-1}\mathbf{C}_t^\top + \mathbf{R}_t, \qquad \text{innovation } \widetilde{\mathbf{y}}_t \;=\; \mathbf{y}_t - \mathbf{C}_t\widehat{\mathbf{x}}_{t|t-1}. \tag{5}$$

Update (Measurement Correction) by Completing the Square. Bayes rule gives

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) \;\propto\; \underbrace{\mathcal{N}(\mathbf{x}_t; \widehat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})}_{\text{prior}} \cdot \underbrace{\mathcal{N}(\mathbf{y}_t; \mathbf{C}_t\mathbf{x}_t, \mathbf{R}_t)}_{\text{likelihood}}.$$

The negative log-density (up to a constant) is the quadratic form

$$\begin{aligned}\mathcal{Q}(\mathbf{x}) &= (\mathbf{x} - \widehat{\mathbf{x}}_{t|t-1})^\top \mathbf{P}_{t|t-1}^{-1}(\mathbf{x} - \widehat{\mathbf{x}}_{t|t-1}) + (\mathbf{y}_t - \mathbf{C}_t\mathbf{x})^\top \mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{C}_t\mathbf{x}) \\ &= \mathbf{x}^\top \Big(\mathbf{P}_{t|t-1}^{-1} + \mathbf{C}_t^\top\mathbf{R}_t^{-1}\mathbf{C}_t\Big)\mathbf{x} - 2\,\mathbf{x}^\top \Big(\mathbf{P}_{t|t-1}^{-1}\widehat{\mathbf{x}}_{t|t-1} + \mathbf{C}_t^\top\mathbf{R}_t^{-1}\mathbf{y}_t\Big) + \text{const.}\end{aligned} \tag{6}$$

Hence the posterior is Gaussian with precision

$$\mathbf{P}_{t|t}^{-1} \;=\; \mathbf{P}_{t|t-1}^{-1} + \mathbf{C}_t^\top\mathbf{R}_t^{-1}\mathbf{C}_t, \qquad \mathbf{P}_{t|t} \;=\; \Big(\mathbf{P}_{t|t-1}^{-1} + \mathbf{C}_t^\top\mathbf{R}_t^{-1}\mathbf{C}_t\Big)^{-1}, \tag{7}$$

and mean

$$\widehat{\mathbf{x}}_{t|t} \;=\; \mathbf{P}_{t|t}\Big(\mathbf{P}_{t|t-1}^{-1}\widehat{\mathbf{x}}_{t|t-1} + \mathbf{C}_t^\top\mathbf{R}_t^{-1}\mathbf{y}_t\Big). \tag{8}$$

Applying the Woodbury identity to Eqn. (7) yields the familiar *Kalman gain* form

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{C}_t^\top \left(\mathbf{C}_t\mathbf{P}_{t|t-1}\mathbf{C}_t^\top + \mathbf{R}_t\right)^{-1} \mathbf{C}_t\mathbf{P}_{t|t-1}. \tag{9}$$

Define

$$\mathbf{K}_t \;\triangleq\; \mathbf{P}_{t|t-1}\mathbf{C}_t^\top\mathbf{S}_t^{-1} \qquad \text{(Kalman gain)}, \tag{10}$$

then substituting $\mathbf{y}_t = \mathbf{C}_t\widehat{\mathbf{x}}_{t|t-1} + \widetilde{\mathbf{y}}_t$ into Eqn. (8) gives the affine update

$$\widehat{\mathbf{x}}_{t|t} = \widehat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\,\widetilde{\mathbf{y}}_t, \tag{11}$$

$$\mathbf{P}_{t|t} = (\mathbf{I}_n - \mathbf{K}_t\mathbf{C}_t)\,\mathbf{P}_{t|t-1}. \tag{12}$$

For numerical symmetry/PSD robustness, use the *Joseph form* (algebraically equivalent to Eqn. (12)):

$$\mathbf{P}_{t|t} \;=\; (\mathbf{I} - \mathbf{K}_t\mathbf{C}_t)\,\mathbf{P}_{t|t-1}\,(\mathbf{I} - \mathbf{K}_t\mathbf{C}_t)^\top + \mathbf{K}_t\mathbf{R}_t\mathbf{K}_t^\top. \tag{13}$$

Proof (Joseph form). *Proof.* Insert Eqn. (10) into Eqn. (13) and expand using $\mathbf{S}_t = \mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}_t^\top + \mathbf{R}_t$ to obtain $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{C}_t^\top \mathbf{S}_t^{-1} \mathbf{C}_t \mathbf{P}_{t|t-1}$, which equals Eqn. (9). ■

1.5 Kalman Gain via MMSE Optimality (Orthogonality Principle)

Proof. Let $\widehat{\mathbf{x}}_{t|t} = \widehat{\mathbf{x}}_{t|t-1} + \mathbf{K}\,\widetilde{\mathbf{y}}_t$ for a matrix $\mathbf{K}$ to be chosen. The posterior error is $\mathbf{e}_t = \mathbf{x}_t - \widehat{\mathbf{x}}_{t|t} = (\mathbf{I} - \mathbf{K}\mathbf{C}_t)\mathbf{e}_t^- - \mathbf{K}\mathbf{v}_t$, where $\mathbf{e}_t^- = \mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}$ is the prediction error ($\mathbb{E}[\mathbf{e}_t^-] = \mathbf{0}$, $\mathbb{E}[\mathbf{e}_t^- \mathbf{e}_t^{-\top}] = \mathbf{P}_{t|t-1}$). Thus

$$\mathbf{P}(\mathbf{K}) \;=\; \mathbb{E}[\mathbf{e}_t \mathbf{e}_t^\top] = (\mathbf{I} - \mathbf{K}\mathbf{C}_t)\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}\mathbf{C}_t)^\top + \mathbf{K}\mathbf{R}_t\mathbf{K}^\top. \tag{14}$$

Minimize $\mathrm{tr}(\mathbf{P}(\mathbf{K}))$ over $\mathbf{K}$. Using matrix calculus ($\partial\, \mathrm{tr}[\mathbf{K}\mathbf{C}\mathbf{P}] = \mathbf{P}^\top \mathbf{C}^\top : \partial \mathbf{K}$, etc.) and setting the derivative to zero yields

$$-\mathbf{C}_t \mathbf{P}_{t|t-1} + \mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}_t^\top \mathbf{K}^\top + \mathbf{R}_t \mathbf{K}^\top = \mathbf{0} \implies \mathbf{K} = \mathbf{P}_{t|t-1} \mathbf{C}_t^\top (\mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}_t^\top + \mathbf{R}_t)^{-1},$$

which is exactly Eqn. (10). Hence the KF update is the unique linear MMSE estimator satisfying the orthogonality principle $\mathbb{E}[\mathbf{e}_t\, \widetilde{\mathbf{y}}_t^\top] = \mathbf{0}$. ■

1.6 Extended Kalman Filter (Nonlinear Dynamics/Observations)

Consider the nonlinear state-space model with additive Gaussian noises:

$$\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{u}_t) + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \tag{15}$$

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t) + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t). \tag{16}$$

Because the posterior is no longer Gaussian in general, the EKF applies a first-order Taylor approximation about the current estimates to obtain a local linear-Gaussian model.

Linearization (First-Order Taylor). Let $\widehat{\mathbf{x}}_{t-1|t-1}$ be the current state estimate. Define the Jacobians

$$\mathbf{F}_t \;\triangleq\; \frac{\partial \mathbf{f}_t}{\partial \mathbf{x}}\bigg|_{\widehat{\mathbf{x}}_{t-1|t-1},\, \mathbf{u}_t} \in \mathbb{R}^{n\times n}, \qquad \mathbf{H}_t \;\triangleq\; \frac{\partial \mathbf{h}_t}{\partial \mathbf{x}}\bigg|_{\widehat{\mathbf{x}}_{t|t-1}} \in \mathbb{R}^{p\times n}.$$

Prediction linearization:

$$\mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{u}_t) \approx \mathbf{f}_t(\widehat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t) + \mathbf{F}_t\, (\mathbf{x}_{t-1} - \widehat{\mathbf{x}}_{t-1|t-1}).$$

Measurement linearization at the predicted mean $\widehat{\mathbf{x}}_{t|t-1}$:

$$\mathbf{h}_t(\mathbf{x}_t) \approx \mathbf{h}_t(\widehat{\mathbf{x}}_{t|t-1}) + \mathbf{H}_t\, (\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}).$$

EKF Recursions. *Prediction:*

$$\widehat{\mathbf{x}}_{t|t-1} = \mathbf{f}_t(\widehat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t), \tag{17}$$

$$\mathbf{P}_{t|t-1} = \mathbf{F}_t\, \mathbf{P}_{t-1|t-1}\, \mathbf{F}_t^\top + \mathbf{Q}_t. \tag{18}$$

Update:

$$\widetilde{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{h}_t(\widehat{\mathbf{x}}_{t|t-1}), \qquad \mathbf{S}_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\top + \mathbf{R}_t, \tag{19}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t^\top \mathbf{S}_t^{-1}, \tag{20}$$

$$\widehat{\mathbf{x}}_{t|t} = \widehat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \widetilde{\mathbf{y}}_t, \tag{21}$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t)\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}_t \mathbf{H}_t)^\top + \mathbf{K}_t \mathbf{R}_t \mathbf{K}_t^\top. \tag{22}$$

Thus EKF is the KF applied to the *locally linearized* model around $(\widehat{\mathbf{x}}_{t-1|t-1}, \widehat{\mathbf{x}}_{t|t-1})$.

Remark (Gauss–Newton view). For Gaussian noises, maximizing the posterior is equivalent to minimizing a nonlinear least-squares cost $\sum_t \|\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{u}_t)\|^2_{\mathbf{Q}_t^{-1}} + \sum_t \|\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t)\|^2_{\mathbf{R}_t^{-1}}$. Linearizing the residuals and taking one Gauss–Newton step at each time yields the EKF update.

1.7 Algorithm (KF and EKF)

Kalman Filter (linear-Gaussian).

1. **Input:** $\{\mathbf{A}_t, \mathbf{B}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t, \mathbf{u}_t\}_{t=1}^T$; prior $(\widehat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$.
2. **For** $t = 1, \ldots, T$**:**
   (a) Predict: Eqns. (3)–(4).
   (b) Innovation: Eqn. (5).
   (c) Gain: Eqn. (10).
   (d) Update: Eqn. (11) and Eqn. (13).
3. **Output:** filtered estimates $\{(\widehat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})\}$.

Extended Kalman Filter (nonlinear).

1. **Input:** $\{\mathbf{f}_t, \mathbf{h}_t, \mathbf{Q}_t, \mathbf{R}_t, \mathbf{u}_t\}_{t=1}^T$; prior $(\widehat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$.
2. **For** $t = 1, \ldots, T$**:**
   (a) Predict mean/cov: Eqns. (17)–(18) using $\mathbf{F}_t = \partial \mathbf{f}_t / \partial \mathbf{x}$ at $(\widehat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t)$.
   (b) Linearize measurement: $\mathbf{H}_t = \partial \mathbf{h}_t / \partial \mathbf{x}$ at $\widehat{\mathbf{x}}_{t|t-1}$.
   (c) Innovation/gain/update: Eqns. (19)–(22).
3. **Output:** EKF estimates $\{(\widehat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})\}$.

1.8 Properties, Optimality, and Stability

- **Optimality (KF).** Under Eqns. (1)–(2) and Gaussian priors, the KF computes the exact posterior mean and covariance; the posterior mean is the MMSE and MAP estimator (they coincide for Gaussians).
- **Innovation whiteness.** $\widetilde{\mathbf{y}}_t$ is zero-mean with covariance $\mathbf{S}_t$ and is uncorrelated with $\widehat{\mathbf{x}}_{t|t}$; this is equivalent to the orthogonality principle.
- **Numerics.** Use Eqn. (13) to preserve symmetry/PSD. Square-root filters propagate Cholesky factors to improve stability.
- **EKF accuracy.** EKF is locally consistent when linearization errors are small; performance can degrade for strong nonlinearities or poor linearization points (use UKF or iterative EKF as alternatives).

1.9 Summary of Variables and Their Dimensions

- $\mathbf{x}_t \in \mathbb{R}^n$: latent state; $\mathbf{y}_t \in \mathbb{R}^p$: measurement; $\mathbf{u}_t \in \mathbb{R}^m$: control.
- $\mathbf{A}_t \in \mathbb{R}^{n\times n}$, $\mathbf{B}_t \in \mathbb{R}^{n\times m}$, $\mathbf{C}_t \in \mathbb{R}^{p\times n}$: linear model matrices.
- $\mathbf{Q}_t \in \mathbb{R}^{n\times n}$, $\mathbf{R}_t \in \mathbb{R}^{p\times p}$: process/measurement covariances ($\succeq \mathbf{0}$).
- $\widehat{\mathbf{x}}_{t|s} \in \mathbb{R}^n$: state estimate; $\mathbf{P}_{t|s} \in \mathbb{R}^{n\times n}$: error covariance.

- $\widetilde{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{C}_t\widehat{\mathbf{x}}_{t|t-1} \in \mathbb{R}^p$: innovation; $\mathbf{S}_t \in \mathbb{R}^{p\times p}$: innovation covariance.
- $\mathbf{K}_t \in \mathbb{R}^{n\times p}$: Kalman gain.
- EKF Jacobians: $\mathbf{F}_t = \partial\mathbf{f}_t/\partial\mathbf{x} \in \mathbb{R}^{n\times n}$, $\mathbf{H}_t = \partial\mathbf{h}_t/\partial\mathbf{x} \in \mathbb{R}^{p\times n}$.

1.10 Summary

From first principles: (i) linear-Gaussian dynamics/observations yield Gaussian priors and likelihoods whose product gives Gaussian posteriors; completing the square produces the KF updates Eqns. (3)–(13) with gain Eqn. (10); (ii) equivalently, the gain is the unique linear MMSE choice minimizing the posterior covariance trace; (iii) EKF applies KF to first-order linearizations of $\mathbf{f}_t, \mathbf{h}_t$, producing Eqns. (17)–(22). These recursions implement an efficient Bayesian estimator with $O(n^3)$ per step dominated by solving $\mathbf{S}_t^{-1}$ and are foundational to tracking, navigation, and sensor fusion.