Extended Kalman Filtering Equations

Bayesian State Estimation in Dynamical Systems

Sai Sampath Kedari
sampath@umich.edu

Contents

1 The Extended Kalman Filter as an Approximate Gaussian Filter

In this section, we derive the Extended Kalman Filter (EKF) as a first-order approximation of the Gaussian filtering equations for nonlinear state-space models. The EKF is obtained by approximating the intractable Gaussian filtering integrals via local Taylor linearizations of the nonlinear dynamics and measurement functions.

Unlike the Kalman filter, where the Gaussian filtering equations reduce to exact, closed-form recursions due to linearity of the dynamics and measurement models, the Extended Kalman Filter is an approximation. The approximation arises because Gaussian distributions are not closed under nonlinear transformations. As a result, the Gaussian filtering integrals cannot be evaluated in closed form and must be approximated.

1.1 Nonlinear Gaussian State-Space Model

We consider the nonlinear Gaussian state-space model

$$X_{k+1} = \Phi(X_k) + \xi_k, \qquad \xi_k \sim \mathcal{N}(0, \Sigma),$$

$$Y_k = h(X_k) + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, \Gamma),$$

where $\Phi(\cdot)$ and $h(\cdot)$ are (possibly) nonlinear functions. We assume $\{\xi_k\}$ and $\{\eta_k\}$ are mutually independent, independent across time, and independent of the state.

1.2 Gaussian Filtering Equations

Gaussian filtering enforces Gaussian predictive and filtering distributions:

$$\mathbb{P}(X_k \mid \mathcal{Y}_{k-1}) \approx \mathcal{N}(m_k^-, C_k^-), \qquad \mathbb{P}(X_k \mid \mathcal{Y}_k) \approx \mathcal{N}(m_k, C_k).$$

The Gaussian filtering recursion is defined by the following moment integrals.

Prediction:

$$m_k^- = \int \Phi(X_{k-1}) \, \mathcal{N}(X_{k-1}; m_{k-1}, C_{k-1}) \, dX_{k-1},$$

$$C_k^- = \int \big(\Phi(X_{k-1}) - m_k^-\big)\big(\Phi(X_{k-1}) - m_k^-\big)^T \mathcal{N}(X_{k-1}; m_{k-1}, C_{k-1}) \, dX_{k-1} + \Sigma.$$

Update:

$$\mu = \int h(X_k) \, \mathcal{N}(X_k; m_k^-, C_k^-) \, dX_k,$$

$$U = \int (X_k - m_k^-)\big(h(X_k) - \mu\big)^T \mathcal{N}(X_k; m_k^-, C_k^-) \, dX_k,$$

$$S = \int \big(h(X_k) - \mu\big)\big(h(X_k) - \mu\big)^T \mathcal{N}(X_k; m_k^-, C_k^-) \, dX_k + \Gamma.$$

The Gaussian filtering update is

$$m_k = m_k^- + US^{-1}(y_k - \mu), \qquad C_k = C_k^- - US^{-1}U^T.$$

1.3 Why Gaussianity Is Not Preserved Under Nonlinear Dynamics

Gaussian distributions are closed under affine transformations, but not under general nonlinear mappings. This is the basic reason Gaussian filtering becomes an approximation for nonlinear systems.

Let

$$X_k \sim \mathcal{N}(m_k, C_k),$$

and consider the nonlinear state transition

$$X_{k+1} = \Phi(X_k) + \xi_k, \qquad \xi_k \sim \mathcal{N}(0, \Sigma).$$

If $\Phi(\cdot)$ is affine, then X_{k+1} is a linear transformation of a Gaussian random variable plus Gaussian noise, hence X_{k+1} is exactly Gaussian. If $\Phi(\cdot)$ is nonlinear, the mapping $X_k \mapsto \Phi(X_k)$ warps the Gaussian probability mass in a nonlinear manner, introducing higher order statistical structure. The resulting distribution is generally non-Gaussian (it may exhibit skewness, heavy tails, or even multimodality), and the additive Gaussian noise smooths but does not restore exact Gaussianity.

The same mechanism applies to the measurement model. Even if X_k is Gaussian, the random variable $h(X_k)$ is Gaussian if and only if $h(\cdot)$ is affine. Therefore the exact Bayesian filtering recursion does not preserve Gaussianity whenever either the dynamics or measurement model is nonlinear. Gaussian filtering methods proceed by approximating the true posterior by a Gaussian distribution characterized only by its first two moments.

1.4 Local Linearization of the Integrands

The EKF approximates the Gaussian filtering integrals by first-order Taylor expansions of $\Phi(\cdot)$ and $h(\cdot)$ around the relevant Gaussian means.

Define the Jacobians

$$A_k = \nabla\Phi(X_{k-1})\big|_{X_{k-1}=m_{k-1}}, \qquad H_k = \nabla h(X_k)\big|_{X_k=m_k^-}.$$

The matrices A_k and H_k define local linear approximations of the nonlinear state transition and measurement functions about the current Gaussian means. All EKF updates are therefore exact Bayesian updates for these locally linearized models, not for the original nonlinear system.

Consequently, EKF inconsistency and divergence arise when the local linear approximation fails to capture the true nonlinear geometry of the state distribution.

First-order Taylor expansions give

$$\Phi(X_{k-1}) \approx \Phi(m_{k-1}) + A_k(X_{k-1} - m_{k-1}),$$

$$h(X_k) \approx h(m_k^-) + H_k(X_k - m_k^-),$$

where higher order terms are neglected.

1.5 Prediction Step

1.5.1 Predicted Mean

Starting from the Gaussian filtering mean integral,

$$m_k^- = \int \Phi(X_{k-1})\,\mathcal{N}(X_{k-1}; m_{k-1}, C_{k-1})\,dX_{k-1},$$

we rewrite it as an expectation under the Gaussian distribution $X_{k-1} \mid \mathcal{Y}_{k-1} \sim \mathcal{N}(m_{k-1}, C_{k-1})$:

$$m_k^- = \mathbb{E}\big[\Phi(X_{k-1}) \mid \mathcal{Y}_{k-1}\big].$$

Applying the first-order Taylor approximation of $\Phi(\cdot)$ about m_{k-1},

$$\Phi(X_{k-1}) \approx \Phi(m_{k-1}) + A_k(X_{k-1} - m_{k-1}),$$

we obtain

$$\begin{aligned} m_k^- &= \mathbb{E}\big[\Phi(X_{k-1}) \mid \mathcal{Y}_{k-1}\big] \\ &\approx \mathbb{E}\big[\Phi(m_{k-1}) + A_k(X_{k-1} - m_{k-1}) \mid \mathcal{Y}_{k-1}\big] \\ &= \mathbb{E}\big[\Phi(m_{k-1}) \mid \mathcal{Y}_{k-1}\big] + \mathbb{E}\big[A_k(X_{k-1} - m_{k-1}) \mid \mathcal{Y}_{k-1}\big] \\ &= \Phi(m_{k-1}) + A_k\,\mathbb{E}\big[X_{k-1} - m_{k-1} \mid \mathcal{Y}_{k-1}\big] \\ &= \Phi(m_{k-1}), \end{aligned}$$

since $\mathbb{E}[X_{k-1} \mid \mathcal{Y}_{k-1}] = m_{k-1}$ implies $\mathbb{E}[X_{k-1} - m_{k-1} \mid \mathcal{Y}_{k-1}] = 0$.

1.5.2 Predicted Covariance

Starting from the Gaussian filtering covariance integral,

$$C_k^- = \int \big(\Phi(X_{k-1}) - m_k^-\big)\big(\Phi(X_{k-1}) - m_k^-\big)^T\,\mathcal{N}(X_{k-1}; m_{k-1}, C_{k-1})\,dX_{k-1} + \Sigma,$$

we rewrite it as

$$C_k^- = \mathbb{E}\Big[\big(\Phi(X_{k-1}) - m_k^-\big)\big(\Phi(X_{k-1}) - m_k^-\big)^T \;\Big|\; \mathcal{Y}_{k-1}\Big] + \Sigma.$$

Using $m_k^- \approx \Phi(m_{k-1})$ and the linearization $\Phi(X_{k-1}) \approx \Phi(m_{k-1}) + A_k(X_{k-1} - m_{k-1})$, we have

$$\Phi(X_{k-1}) - m_k^- \approx \big(\Phi(m_{k-1}) + A_k(X_{k-1} - m_{k-1})\big) - \Phi(m_{k-1}) = A_k(X_{k-1} - m_{k-1}).$$

Substituting into the expectation,

$$\begin{aligned} C_k^- &\approx \mathbb{E}\Big[A_k(X_{k-1} - m_{k-1})\big(A_k(X_{k-1} - m_{k-1})\big)^T \;\Big|\; \mathcal{Y}_{k-1}\Big] + \Sigma \\ &= \mathbb{E}\Big[A_k(X_{k-1} - m_{k-1})(X_{k-1} - m_{k-1})^T A_k^T \;\Big|\; \mathcal{Y}_{k-1}\Big] + \Sigma \\ &= A_k\,\mathbb{E}\Big[(X_{k-1} - m_{k-1})(X_{k-1} - m_{k-1})^T \;\Big|\; \mathcal{Y}_{k-1}\Big]\,A_k^T + \Sigma \\ &= A_k C_{k-1} A_k^T + \Sigma. \end{aligned}$$

1.6 Update Step

Assume the predictive distribution

$$X_k \mid \mathcal{Y}_{k-1} \sim \mathcal{N}(m_k^-, C_k^-).$$

We approximate the update integrals by linearizing $h(\cdot)$ around m_k^-:

$$h(X_k) \approx h(m_k^-) + H_k(X_k - m_k^-).$$

1.6.1 Predicted Measurement Mean

From the Gaussian filtering definition,

$$\mu = \int h(X_k)\,\mathcal{N}(X_k; m_k^-, C_k^-)\,dX_k = \mathbb{E}\big[h(X_k) \mid \mathcal{Y}_{k-1}\big].$$

Using the linearization,

$$\begin{aligned}
\mu &= \mathbb{E}\big[h(X_k) \mid \mathcal{Y}_{k-1}\big] \\
&\approx \mathbb{E}\big[h(m_k^-) + H_k(X_k - m_k^-) \mid \mathcal{Y}_{k-1}\big] \\
&= h(m_k^-) + H_k\,\mathbb{E}\big[X_k - m_k^- \mid \mathcal{Y}_{k-1}\big] \\
&= h(m_k^-),
\end{aligned}$$

since $\mathbb{E}[X_k \mid \mathcal{Y}_{k-1}] = m_k^-$.

1.6.2 Cross-Covariance

From the Gaussian filtering definition,

$$U = \int (X_k - m_k^-)\big(h(X_k) - \mu\big)^T \mathcal{N}(X_k; m_k^-, C_k^-)\,dX_k.$$

Using $\mu \approx h(m_k^-)$ and the linearization of $h(X_k)$,

$$h(X_k) - \mu \approx \big(h(m_k^-) + H_k(X_k - m_k^-)\big) - h(m_k^-) = H_k(X_k - m_k^-).$$

Therefore,

$$\begin{aligned}
U &\approx \mathbb{E}\Big[(X_k - m_k^-)\big(H_k(X_k - m_k^-)\big)^T \,\Big|\, \mathcal{Y}_{k-1}\Big] \\
&= \mathbb{E}\Big[(X_k - m_k^-)(X_k - m_k^-)^T H_k^T \,\Big|\, \mathcal{Y}_{k-1}\Big] \\
&= \mathbb{E}\Big[(X_k - m_k^-)(X_k - m_k^-)^T \,\Big|\, \mathcal{Y}_{k-1}\Big] H_k^T \\
&= C_k^- H_k^T.
\end{aligned}$$

1.6.3 Innovation Covariance

From the Gaussian filtering definition,

$$S = \int \big(h(X_k) - \mu\big)\big(h(X_k) - \mu\big)^T \mathcal{N}(X_k; m_k^-, C_k^-)\,dX_k + \Gamma.$$

Using $h(X_k) - \mu \approx H_k(X_k - m_k^-)$, we obtain

$$\begin{aligned}
S &\approx \mathbb{E}\Big[\big(H_k(X_k - m_k^-)\big)\big(H_k(X_k - m_k^-)\big)^T \,\Big|\, \mathcal{Y}_{k-1}\Big] + \Gamma \\
&= \mathbb{E}\Big[H_k(X_k - m_k^-)(X_k - m_k^-)^T H_k^T \,\Big|\, \mathcal{Y}_{k-1}\Big] + \Gamma \\
&= H_k\,\mathbb{E}\Big[(X_k - m_k^-)(X_k - m_k^-)^T \,\Big|\, \mathcal{Y}_{k-1}\Big]\, H_k^T + \Gamma \\
&= H_k C_k^- H_k^T + \Gamma.
\end{aligned}$$

1.7 Extended Kalman Filtering Equations (EKF)

Substituting the locally linearized expressions into the Gaussian filtering recursion yields the Extended Kalman Filter. The resulting equations have the same algebraic structure as the Kalman filter, but with time-varying Jacobians computed along the estimated trajectory.

Prediction Step

$$A_k = \nabla\Phi(X_{k-1})\big|_{X_{k-1}=m_{k-1}}$$

$$m_k^- = \Phi(m_{k-1}),$$

$$C_k^- = A_k C_{k-1} A_k^T + \Sigma.$$

Update Step

$$H_k = \nabla h(X_k)\big|_{X_k=m_k^-}$$

$$\mu = h(m_k^-),$$

$$U = C_k^- H_k^T,$$

$$S = H_k C_k^- H_k^T + \Gamma.$$

The filtering update is then

$$m_k = m_k^- + US^{-1}(y_k - \mu),$$

$$C_k = C_k^- - US^{-1}U^T.$$

1.8 Conclusion

The Extended Kalman Filter is therefore a Gaussian filter in which the required moments are approximated using first-order Taylor expansions. The quality of the approximation depends critically on the degree of nonlinearity and the local validity of the linearization.

In subsequent chapters, we derive alternative Gaussian filtering methods (Unscented Kalman Filter, Gauss–Hermite filtering, and cubature filters) that approximate these same Gaussian filtering integrals without relying on local linearization.