Gaussian Filtering in Bayesian State Estimation

Bayesian State Estimation in Dynamical Systems

Sai Sampath Kedari
sampath@umich.edu

Contents

1 Gaussian Filtering: Motivation and Overview

In the previous report, we derived the *exact Bayesian filtering equations* for general discrete-time state-space models. These equations provide a principled recursive framework for state estimation under uncertainty and apply to *arbitrary* dynamical systems.

Importantly, the derivation of the Bayesian filtering equations is completely generic. It does not assume linear dynamics, linear measurement models, Gaussian noise, or any specific parametric form of the state or observation distributions. The derivation relies only on the Markov property of the state evolution and on the conditional independence structure implied by the state-space model.

As a result, the Bayesian filtering recursion applies equally to linear and nonlinear systems, and to systems with Gaussian or non-Gaussian process and measurement noise. However, while the filtering equations are exact, their practical evaluation is generally intractable for nonlinear systems, since the required probability distributions quickly become analytically and computationally intractable.

Gaussian filtering addresses this challenge by introducing a structured approximation that enables tractable recursive inference. Specifically, it approximates the exact Bayesian filtering recursion by restricting all intermediate belief distributions to be Gaussian and by propagating only their first two moments.

From a mathematical perspective, Gaussian filtering replaces the exact integrals defining the predictive and filtering distributions with approximations that match only the first two moments of those distributions. This distinction is important, as different Gaussian filtering methods differ precisely in how these moments are approximated.

This report is devoted to the derivation and interpretation of the *Gaussian filtering equations*, which arise from applying this Gaussian approximation to the exact Bayesian filtering framework derived previously.

1.1 State-Space Model

We consider the nonlinear state-space model

$$\begin{aligned} X_{k+1} &= \Phi(X_k;\theta_1) + \xi_k, \qquad && \xi_k \sim \mathcal{N}(0,\Sigma), \\ Y_k &= h(X_k;\theta_2) + \eta_k, \qquad && \eta_k \sim \mathcal{N}(0,\Gamma), \end{aligned}$$

where

- X_k denotes the latent system state,
- Y_k denotes the measurement at time k,
- $\Phi(\cdot)$ is the (possibly nonlinear) dynamics function,
- $h(\cdot)$ is the (possibly nonlinear) measurement function,
- ξ_k and η_k are zero-mean Gaussian process and measurement noises, respectively.

The assumption of Gaussian process and measurement noise ensures that the conditional transition and likelihood models are Gaussian, making moment-based approximations well-defined within the Gaussian filtering framework.

1.2 Assumptions on Parameters

Throughout this report, we focus exclusively on *state estimation*. The model parameters θ_1 and θ_2 are assumed to be either known *a priori* or treated as random variables that have been marginalized out. In either case, these parameters do not enter explicitly into the filtering equations derived here.

Parameter estimation and learning are therefore outside the scope of this report; the objective is solely to estimate the latent system state given noisy measurements.

1.3 Why Gaussian Filtering Is an Approximation

A key point that must be emphasized is that Gaussianity is not preserved under general nonlinear transformations.

Suppose the initial state is Gaussian:

$$X_0 \sim \mathcal{N}(m_0, C_0).$$

The next state is given by

$$X_1 = \Phi(X_0; \theta_1) + \xi_0, \qquad \xi_0 \sim \mathcal{N}(0, \Sigma).$$

The random variable X_1 is Gaussian if and only if $\Phi(X_0; \theta_1)$ is an affine function of X_0, that is,

$$\Phi(X_0; \theta_1) = A(\theta_1)X_0 + b(\theta_1).$$

In this case, X_1 is a linear transformation of a Gaussian random variable plus Gaussian noise, and the resulting distribution is exactly Gaussian.

If $\Phi(\cdot)$ is nonlinear, then $\Phi(X_0; \theta_1)$ is a nonlinear transformation of a Gaussian random variable, and the resulting distribution of X_1 is, in general, non-Gaussian. An analogous argument applies to the measurement model: even if X_k is Gaussian, the measurement Y_k is Gaussian if and only if $h(\cdot)$ is an affine function.

Therefore, for nonlinear dynamics or nonlinear measurement models, the exact Bayesian filtering recursion produces non-Gaussian distributions at each time step.

Gaussian filtering is explicitly an *approximation method*. At each time step, it replaces the true, generally non-Gaussian, distributions by Gaussian distributions whose mean and covariance match the first two moments of the exact distribution.

2 Gaussian Filtering Equations

Gaussian filtering approximates the exact Bayesian filtering recursion by restricting the predictive and filtering distributions to be Gaussian at every time step. This approximation replaces the exact distribution-valued filtering equations with recursive equations for the mean and covariance of the state.

Specifically, Gaussian filtering enforces

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

The Gaussian filtering equations are obtained by approximating the exact Bayesian filtering integrals through moment matching.

2.1 Gaussian Filtering Recursion

The Gaussian filtering recursion consists of a prediction step followed by an update step.

Prediction: The predicted mean and covariance are

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

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

Update: Define

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

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

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

The updated mean and covariance are

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

2.2 Gaussian Prediction: Mean and Covariance Derivations

We start from the exact prediction equation

$$\mathbb{P}(X_k \mid \mathcal{Y}_{k-1}) = \int \mathbb{P}(X_k \mid X_{k-1})\, \mathbb{P}(X_{k-1} \mid \mathcal{Y}_{k-1})\, dX_{k-1},$$

and assume the Gaussian filtering approximation at time $k-1$

$$\mathbb{P}(X_{k-1} \mid \mathcal{Y}_{k-1}) \approx \mathcal{N}(X_{k-1}; m_{k-1}, C_{k-1}),$$

with dynamics

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

so that

$$\mathbb{P}(X_k \mid X_{k-1}) = \mathcal{N}\big(X_k;\, \Phi(X_{k-1}), \Sigma\big).$$

Notational abuse (Gaussian-weighted integrals). For a measurable function g and Gaussian weight $\mathcal{N}(x; m, C)$, we use

$$\int g(x)\, \mathcal{N}(x; m, C)\, dx \;\equiv\; \mathbb{E}[g(X)], \qquad X \sim \mathcal{N}(m, C).$$

2.2.1 Predicted Mean

By definition,

$$m_k^- \;\equiv\; \mathbb{E}[X_k \mid \mathcal{Y}_{k-1}] = \int x_k \, \mathbb{P}(x_k \mid \mathcal{Y}_{k-1}) \, dx_k.$$

Substitute the prediction equation:

$$m_k^- = \int x_k \left[\int \mathbb{P}(x_k \mid x_{k-1}) \, \mathbb{P}(x_{k-1} \mid \mathcal{Y}_{k-1}) \, dx_{k-1} \right] dx_k.$$

Interchange integrals (Fubini):

$$m_k^- = \int \left[\int x_k \, \mathbb{P}(x_k \mid x_{k-1}) \, dx_k \right] \mathbb{P}(x_{k-1} \mid \mathcal{Y}_{k-1}) \, dx_{k-1}.$$

The inner integral is the conditional mean under $\mathbb{P}(x_k \mid x_{k-1}) = \mathcal{N}(x_k; \Phi(x_{k-1}), \Sigma)$:

$$\int x_k \, \mathbb{P}(x_k \mid x_{k-1}) \, dx_k = \mathbb{E}[X_k \mid x_{k-1}] = \Phi(x_{k-1}),$$

since $\mathbb{E}[\xi_{k-1}] = 0$. Hence

$$m_k^- = \int \Phi(x_{k-1}) \, \mathbb{P}(x_{k-1} \mid \mathcal{Y}_{k-1}) \, dx_{k-1} \approx \int \Phi(x_{k-1}) \, \mathcal{N}(x_{k-1}; m_{k-1}, C_{k-1}) \, dx_{k-1}.$$

2.2.2 Predicted Covariance

Define the predictive covariance

$$C_k^- \;\equiv\; \mathrm{Cov}[X_k \mid \mathcal{Y}_{k-1}] = \mathbb{E}\big[(X_k - m_k^-)(X_k - m_k^-)^T \mid \mathcal{Y}_{k-1}\big] \,.$$

Using the conditional form $X_k = \Phi(X_{k-1}) + \xi_{k-1}$, we write

$$X_k - m_k^- = \big(\Phi(X_{k-1}) - m_k^-\big) + \xi_{k-1}.$$

Therefore,

$$\begin{aligned}(X_k - m_k^-)(X_k - m_k^-)^T = {}& \big(\Phi(X_{k-1}) - m_k^-\big)\big(\Phi(X_{k-1}) - m_k^-\big)^T \\ &+ \big(\Phi(X_{k-1}) - m_k^-\big)\xi_{k-1}^T + \xi_{k-1}\big(\Phi(X_{k-1}) - m_k^-\big)^T + \xi_{k-1}\xi_{k-1}^T.\end{aligned}$$

Take conditional expectation given X_{k-1}:

$$\mathbb{E}[\xi_{k-1} \mid X_{k-1}] = 0, \qquad \mathbb{E}[\xi_{k-1}\xi_{k-1}^T \mid X_{k-1}] = \Sigma,$$

so the cross terms vanish and we obtain

$$\mathbb{E}\big[(X_k - m_k^-)(X_k - m_k^-)^T \mid X_{k-1}\big] = \big(\Phi(X_{k-1}) - m_k^-\big)\big(\Phi(X_{k-1}) - m_k^-\big)^T + \Sigma.$$

Now take expectation with respect to $\mathbb{P}(X_{k-1} \mid \mathcal{Y}_{k-1})$:

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

2.3 Gaussian Update: Mean and Covariance Derivations

We now derive the Gaussian filtering *update equations*. The derivation proceeds by constructing the joint distribution of the predicted state X_k and the measurement Y_k, and then conditioning on the observed measurement $Y_k = y_k$.

Throughout this section, we assume the Gaussian filtering approximation

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

and the observation model

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

with η_k independent of X_k.

2.3.1 Joint Mean

From the observation model, the conditional expectation of Y_k given X_k is

$$\mathbb{E}[Y_k \mid X_k] = h(X_k).$$

Applying the law of total expectation,

$$\mathbb{E}[Y_k \mid \mathcal{Y}_{k-1}] = \mathbb{E}_{X_k|\mathcal{Y}_{k-1}}[\mathbb{E}[Y_k \mid X_k]] = \mathbb{E}_{X_k|\mathcal{Y}_{k-1}}[h(X_k)].$$

Define

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

The joint mean of the stacked random vector

$$Z_k = \begin{bmatrix} X_k \\ Y_k \end{bmatrix}$$

is therefore

$$\mathbb{E}[Z_k \mid \mathcal{Y}_{k-1}] = \begin{bmatrix} m_k^- \\ \mu \end{bmatrix}.$$

2.3.2 Joint Covariance

The joint covariance matrix admits the block structure

$$\mathrm{Cov}(Z_k \mid \mathcal{Y}_{k-1}) = \begin{bmatrix} \mathrm{Cov}(X_k, X_k) & \mathrm{Cov}(X_k, Y_k) \\ \mathrm{Cov}(Y_k, X_k) & \mathrm{Cov}(Y_k, Y_k) \end{bmatrix}.$$

Each block is computed explicitly below.

State Covariance By definition,

$$\mathrm{Cov}(X_k, X_k \mid \mathcal{Y}_{k-1}) = C_k^-.$$

Cross-Covariance Using covariance algebra,

$$\mathrm{Cov}(X_k, Y_k \mid \mathcal{Y}_{k-1}) = \mathrm{Cov}(X_k, h(X_k) + \eta_k \mid \mathcal{Y}_{k-1}).$$

Expanding the covariance,

$$\mathrm{Cov}(X_k, Y_k) = \mathrm{Cov}(X_k, h(X_k)) + \mathrm{Cov}(X_k, \eta_k).$$

Since X_k and η_k are independent,

$$\mathrm{Cov}(X_k, \eta_k) = 0.$$

Define the cross-covariance

$$U \;\equiv\; \mathrm{Cov}(X_k, h(X_k) \mid \mathcal{Y}_{k-1}) = \int (x_k - m_k^-)\big(h(x_k) - \mu\big)^T \mathcal{N}(x_k; m_k^-, C_k^-)\, dx_k.$$

Thus,

$$\mathrm{Cov}(X_k, Y_k \mid \mathcal{Y}_{k-1}) = U, \qquad \mathrm{Cov}(Y_k, X_k \mid \mathcal{Y}_{k-1}) = U^T.$$

Measurement Covariance To compute $\mathrm{Cov}(Y_k, Y_k \mid \mathcal{Y}_{k-1})$, we use the law of total variance:

$$\mathrm{Var}(Y_k) = \mathbb{E}[\mathrm{Var}(Y_k \mid X_k)] + \mathrm{Var}(\mathbb{E}[Y_k \mid X_k])\,.$$

From the observation model,

$$\mathrm{Var}(Y_k \mid X_k) = \Gamma,$$

and therefore

$$\mathbb{E}[\mathrm{Var}(Y_k \mid X_k)] = \Gamma.$$

Moreover,

$$\mathbb{E}[Y_k \mid X_k] = h(X_k),$$

so

$$\mathrm{Var}(h(X_k)) = \int \big(h(x_k) - \mu\big)\big(h(x_k) - \mu\big)^T \mathcal{N}(x_k; m_k^-, C_k^-)\, dx_k.$$

Define

$$S \;\equiv\; \mathrm{Cov}(Y_k, Y_k \mid \mathcal{Y}_{k-1}) = \int \big(h(x_k) - \mu\big)\big(h(x_k) - \mu\big)^T \mathcal{N}(x_k; m_k^-, C_k^-)\, dx_k + \Gamma.$$

2.3.3 Joint Gaussian Representation

Collecting the results, the joint distribution of (X_k, Y_k) conditioned on $\mathcal{Y}_{k-1}$ is

$$\begin{bmatrix} X_k \\ Y_k \end{bmatrix} \Big|\, \mathcal{Y}_{k-1} \sim \mathcal{N}\left(\begin{bmatrix} m_k^- \\ \mu \end{bmatrix}, \begin{bmatrix} C_k^- & U \\ U^T & S \end{bmatrix} \right).$$

2.3.4 Conditioning on the Measurement

We now condition the joint Gaussian distribution on the observed measurement $Y_k = y_k$.

From the standard conditioning formula for jointly Gaussian random vectors,

$$X_k \mid Y_k = y_k, \mathcal{Y}_{k-1} \sim \mathcal{N}\left(m_k^- + US^{-1}(y_k - \mu),\ C_k^- - US^{-1}U^T\right).$$

Therefore, the Gaussian filtering update equations are

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

3 Outlook: Approximate Gaussian Filtering Algorithms

The Gaussian filtering equations derived in this report define a general moment-based recursive inference framework. For nonlinear dynamical systems, the key computational challenge lies in evaluating the Gaussian-weighted integrals appearing in the prediction and update steps.

Different Gaussian filtering algorithms arise from different approximation strategies for these integrals. The Extended Kalman Filter (EKF) approximates the integrands via local Taylor series expansions, leading to closed-form expressions based on linearization. In contrast, the Gauss–Hermite Kalman Filter (GHKF), Cubature Kalman Filter (CKF), and Unscented Kalman Filter (UKF) approximate the same integrals using numerical quadrature rules of increasing accuracy and computational cost.

In the subsequent sections, we develop these algorithms as concrete instantiations of the Gaussian filtering equations, each corresponding to a specific approximation of the underlying moment integrals.