Bayesian Filtering Equations

Recursive State Estimation in Dynamical Systems

Sai Sampath Kedari
sampath@umich.edu

Contents

1 Filtering Equations

We now specialize the Bayesian modeling framework to the problem of **state estimation** for discrete-time dynamical systems. The objective of filtering is to recursively estimate the latent system state as new measurements arrive.

We consider the same state-space model introduced previously:

$$X_{k+1} = \Phi(X_k; \theta_1) + \xi_k, \qquad \xi_k \sim \mathcal{N}(0, \Sigma),$$

$$Y_k = h(X_k; \theta_2) + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, \Gamma).$$

The system evolves in discrete time, and measurements are acquired at each time step through noisy sensors.

1.1 State Estimation Objective

The central inference task is to estimate the system state from noisy measurements. At time step k, the state X_k is not directly observable and must be inferred from the measurement history.

Ideally, we would like to compute the full posterior distribution over all states up to time n given all available measurements:

$$P(X_0, X_1, \dots, X_n \mid Y_1, Y_2, \dots, Y_n).$$

This joint posterior contains complete probabilistic information about the entire state trajectory.

1.2 Assumptions on Model Parameters

In state estimation, we typically assume that the parameters θ_1 and θ_2 are known. If they are not known, one may treat them as random variables and marginalize them out. In either case, our primary interest in this work is *not* in estimating the parameters themselves, but in estimating the system states.

Therefore, throughout this report, we assume that θ_1 and θ_2 are either known or have been appropriately handled, and we focus solely on state estimation.

1.3 Full Bayesian Posterior

Using Bayes' rule, the full posterior can be written as

$$P(X_{0:n} \mid Y_{1:n}) = \frac{P(Y_{1:n} \mid X_{0:n})\, P(X_{0:n})}{P(Y_{1:n})}.$$

In principle, this posterior can be computed explicitly by repeated application of Bayes' rule as new measurements arrive.

1.4 Why Full Trajectory Inference Is Infeasible

Although the full posterior distribution is theoretically well-defined, explicit computation of

$$P(X_0, X_1, \dots, X_n \mid Y_1, \dots, Y_n)$$

is not feasible in real-time applications.

The reasons are fundamental:

- At each new time step, a new latent state variable is introduced.
- The dimensionality of the posterior distribution grows linearly with time.
- The computational cost per time step increases as new measurements arrive.

As a result, the cost of computing and storing the full posterior becomes intractable for online sensing and control systems.

In real-time robotic applications, such as state estimation for humanoids, quadrupeds, or mobile robots, inference algorithms must operate with bounded computational complexity at every time step.

1.5 Filtering as a Recursive Approximation

To address the challenges associated with full trajectory inference, filtering focuses on estimating only the *current state* rather than the entire state history.

Filtering does not attempt to estimate the entire state trajectory or revise past state estimates; it only maintains a belief over the current state.

At time step k, the filtering objective is to compute

$$P(X_k \mid Y_1, Y_2, \dots, Y_k),$$

the posterior distribution of the current state given all measurements available up to that time.

Filtering exploits the Markov structure of the state-space model to derive recursive update equations whose computational complexity does not grow with time. By maintaining only the belief over the current state, filtering provides a tractable and principled solution for real-time state estimation under uncertainty.

1.6 Scope of Filtering Methods

Filtering methods differ in how the posterior distribution $P(X_k \mid Y_{1:k})$ is represented and propagated.

In general, exact computation of this posterior is intractable for nonlinear systems. This motivates the development of approximate filtering techniques, including:

- Gaussian filtering methods
- Linearization-based filters
- Sampling-based filters

All such methods are derived from the same Bayesian filtering recursion, which forms the foundation of state estimation algorithms used in robotics, control, and learning.

2 Bayesian Filtering Equations

In the previous section, we formulated the Bayesian modeling framework for discrete-time dynamical systems and discussed why full trajectory inference is infeasible in real-time applications. We now turn to the central problem of this report: deriving the **Bayesian filtering equations**.

The objective of filtering is to compute, at each time step n, the posterior distribution of the current state given all measurements available up to that time,

$$P(X_n \mid \mathcal{Y}_n), \qquad \mathcal{Y}_n = \{Y_1, Y_2, \ldots, Y_n\}.$$

A naive approach would be to recompute this posterior from scratch each time a new measurement arrives, using the entire measurement history. However, as discussed earlier, such an approach is computationally intractable because the dimensionality of the joint distribution grows with time.

The key idea behind filtering is to exploit the temporal structure of the state-space model in order to derive a *recursive* update rule. Rather than operating on the entire measurement history, filtering algorithms maintain and update a belief over the current state as new data becomes available.

In this section, we formalize this recursive viewpoint and derive the filtering equations that govern the evolution of the posterior distribution over time. These equations form the foundation of all filtering methods used in robotics, control, and learning.

2.1 Filtering as a Recursive Mapping

The filtering problem can be viewed as the construction of a recursive mapping that updates the belief over the system state as new data arrives.

At time step $n-1$, assume that all measurements up to that time have already been incorporated, and that the system state is described by the posterior distribution

$$P(X_{n-1} \mid \mathcal{Y}_{n-1}),$$

where $\mathcal{Y}_{n-1} = \{Y_1, Y_2, \ldots, Y_{n-1}\}$ denotes the collection of all observations available up to time $n-1$.

At the next time step, a new measurement y_n becomes available. The objective of filtering is to update the belief over the system state using this new information and to compute the posterior distribution

$$P(X_n \mid \mathcal{Y}_n), \qquad \mathcal{Y}_n = \mathcal{Y}_{n-1} \cup \{Y_n\}.$$

Rather than recomputing the posterior distribution from scratch using the entire measurement history, filtering seeks a recursive update rule. Specifically, we aim to determine a mapping

$$g : \Big(P(X_{n-1} \mid \mathcal{Y}_{n-1}),\ y_n\Big) \longmapsto P(X_n \mid \mathcal{Y}_n),$$

which transforms the previous posterior distribution and the new measurement into the current posterior distribution.

The motivation for this recursive formulation is computational. An alternative approach would be to accumulate all measurements $\mathcal{Y}_n$ and compute the full joint posterior over all states $X_{0:n}$, followed by marginalization to obtain $P(X_n \mid \mathcal{Y}_n)$. However, this approach becomes infeasible as the number of time steps grows, since the dimensionality of the joint distribution increases with time.

By contrast, the recursive mapping g leverages the fact that the uncertainty associated with past measurements has already been encoded in the previous posterior distribution $P(X_{n-1} \mid \mathcal{Y}_{n-1})$. This allows the filtering update at time n to depend only on the previous posterior and the new observation, rather than on the entire measurement history.

The central goal of filtering theory is therefore to characterize the mapping g and to determine how it can be evaluated efficiently under the modeling assumptions imposed by the dynamical system and observation models.

2.2 Decomposition of the Recursive Update

To derive a recursive update rule for filtering, we must explicitly separate what is already known from what is newly observed. At time step n, the available information consists of two parts:

- past measurements $\mathcal{Y}_{n-1} = \{Y_1, \dots, Y_{n-1}\}$,
- the new measurement y_n.

Our objective is to compute the posterior distribution

$$\mathbb{P}(X_n \mid \mathcal{Y}_n), \qquad \mathcal{Y}_n = \mathcal{Y}_{n-1} \cup \{Y_n\}.$$

Rather than attempting to compute this distribution directly, we introduce the previous state X_{n-1} as an auxiliary variable. This allows us to construct a joint distribution over consecutive states and then marginalize out the introduced variable. Such a strategy is common in recursive inference algorithms and appears not only in filtering, but also in methods such as importance sampling and accept–reject sampling.

Using marginalization, we write

$$\mathbb{P}(X_n \mid \mathcal{Y}_n) = \int \mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1}, y_n)\, dX_{n-1}.$$

This identity holds generally and does not yet rely on any modeling assumptions.

To proceed, we apply conditional Bayes' rule with respect to the new measurement y_n, while conditioning on the previously available data $\mathcal{Y}_{n-1}$. This yields

$$\mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1}, y_n) = \frac{\mathbb{P}(y_n \mid X_{n-1}, X_n, \mathcal{Y}_{n-1})\, \mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}.$$

Substituting this expression back into the marginalization step gives

$$\mathbb{P}(X_n \mid \mathcal{Y}_n) = \int \frac{\mathbb{P}(y_n \mid X_{n-1}, X_n, \mathcal{Y}_{n-1})\, \mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}\, dX_{n-1}.$$

The denominator $\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})$ acts as a normalization constant, often referred to as the *evidence* or *marginal likelihood*. It is given explicitly by

$$\mathbb{P}(y_n \mid \mathcal{Y}_{n-1}) = \int \int \mathbb{P}(y_n \mid X_{n-1}, X_n, \mathcal{Y}_{n-1})\, \mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1})\, dX_{n-1}\, dX_n.$$

At this stage, the expression for the posterior is exact but not yet useful. To simplify it further, we exploit the conditional independence properties encoded in the Bayesian network of the state-space model.

The derivation relies on the Markov property of the state dynamics and the conditional independence of measurements given the current state.

In particular, the graphical structure implies the following d-separation relationships:

$$Y_n \perp X_{n-1} \mid X_n, \qquad Y_n \perp \mathcal{Y}_{n-1} \mid X_n.$$

As a consequence, the likelihood term simplifies to

$$\mathbb{P}(y_n \mid X_{n-1}, X_n, \mathcal{Y}_{n-1}) = \mathbb{P}(y_n \mid X_n).$$

This conditional independence will play a crucial role in obtaining a tractable recursive filtering equation. The remaining steps of the derivation build on this simplification and will be developed next.

We now substitute the simplified likelihood term into the expression for the posterior. Using the conditional independence result

$$\mathbb{P}(y_n \mid X_{n-1}, X_n, \mathcal{Y}_{n-1}) = \mathbb{P}(y_n \mid X_n),$$

we obtain

$$\mathbb{P}(X_n \mid \mathcal{Y}_n) = \int \frac{\mathbb{P}(y_n \mid X_n)\, \mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}\, dX_{n-1}.$$

The likelihood term $\mathbb{P}(y_n \mid X_n)$ and the evidence term $\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})$ do not depend on X_{n-1}. They can therefore be taken outside the integral, yielding

$$\mathbb{P}(X_n \mid \mathcal{Y}_n) = \frac{\mathbb{P}(y_n \mid X_n)}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})} \int \mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1})\, dX_{n-1}.$$

At this stage, the remaining integral involves the joint distribution $\mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1})$. This joint distribution is not directly available, since the model specifies the dynamics through the conditional distribution $\mathbb{P}(X_n \mid X_{n-1})$, rather than through a joint model over (X_{n-1}, X_n).

To proceed, we apply the conditional probability rule and write

$$\mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1}) = \mathbb{P}(X_n \mid X_{n-1}, \mathcal{Y}_{n-1})\, \mathbb{P}(X_{n-1} \mid \mathcal{Y}_{n-1}).$$

This factorization is exact and follows directly from the definition of conditional probability.

Next, we exploit the conditional independence structure of the Bayesian network. In particular, the Markov property of the state dynamics implies

$$X_n \perp \mathcal{Y}_{n-1} \mid X_{n-1}.$$

As a result, the conditional distribution simplifies to

$$\mathbb{P}(X_n \mid X_{n-1}, \mathcal{Y}_{n-1}) = \mathbb{P}(X_n \mid X_{n-1}).$$

Substituting this into the joint distribution yields

$$\mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1}) = \mathbb{P}(X_n \mid X_{n-1})\, \mathbb{P}(X_{n-1} \mid \mathcal{Y}_{n-1}).$$

Replacing this expression in the posterior gives

$$\mathbb{P}(X_n \mid \mathcal{Y}_n) = \frac{\mathbb{P}(y_n \mid X_n)}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})} \int \mathbb{P}(X_n \mid X_{n-1})\, \mathbb{P}(X_{n-1} \mid \mathcal{Y}_{n-1})\, dX_{n-1}.$$

At this point, each term in the expression admits a clear interpretation:

- $\mathbb{P}(X_n \mid X_{n-1})$ is the dynamics model, which is assumed to be known,
- $\mathbb{P}(y_n \mid X_n)$ is the observation model,
- $\mathbb{P}(X_{n-1} \mid \mathcal{Y}_{n-1})$ is the posterior distribution from the previous time step,
- $\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})$ is the evidence or marginal likelihood, which serves as a normalization constant.

In many inference settings, this normalization constant does not need to be computed explicitly. For example, in conjugate Bayesian models the normalization is available in closed form, while in sampling-based methods such as Markov chain Monte Carlo, only an unnormalized version of the target distribution is required. As a result, the evidence does not affect state estimates and serves only as a normalization constant.

The remaining integral can be written as

$$\int \mathbb{P}(X_n \mid X_{n-1})\, \mathbb{P}(X_{n-1} \mid \mathcal{Y}_{n-1})\, dX_{n-1} = \mathbb{P}(X_n \mid \mathcal{Y}_{n-1}).$$

This equality follows directly from the rules of probability and represents the marginalization of the joint distribution $\mathbb{P}(X_{n-1}, X_n \mid \mathcal{Y}_{n-1})$ with respect to X_{n-1}.

The distribution $\mathbb{P}(X_n \mid \mathcal{Y}_{n-1})$ is referred to as the *predictive distribution*. It represents the belief over the state at time n obtained by propagating the previous posterior $\mathbb{P}(X_{n-1} \mid \mathcal{Y}_{n-1})$ forward through the dynamics model, prior to incorporating the new measurement.

Using this definition, the filtering posterior can be written as

$$\mathbb{P}(X_n \mid \mathcal{Y}_n) = \frac{\mathbb{P}(y_n \mid X_n)}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}\, \mathbb{P}(X_n \mid \mathcal{Y}_{n-1}).$$

This expression makes explicit the two fundamental operations of Bayesian filtering: prediction, in which uncertainty is propagated forward through the dynamics model, and update, in which the predictive distribution is corrected using the new observation.

2.3 Prediction and Update Equations

We now state the Bayesian filtering recursion in its standard prediction–update form. The goal at each time step k is to compute the filtering distribution $\mathbb{P}(X_k \mid \mathcal{Y}_k)$, where $\mathcal{Y}_k = \{Y_1, \ldots, Y_k\}$ denotes the available measurement history.

Initialization The recursion starts from a prior distribution over the initial state:

$$\mathbb{P}(X_0).$$

Prediction Given the filtering distribution at time $k-1$, the predictive distribution at time k is obtained by marginalization (Chapman–Kolmogorov equation):

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

Update After observing the new measurement y_k, the filtering distribution is obtained by Bayes' rule:

$$\mathbb{P}(X_k \mid \mathcal{Y}_k) = \frac{\mathbb{P}(y_k \mid X_k)\, \mathbb{P}(X_k \mid \mathcal{Y}_{k-1})}{\mathbb{P}(y_k \mid \mathcal{Y}_{k-1})}.$$

Equivalently, in unnormalized form,

$$\mathbb{P}(X_k \mid \mathcal{Y}_k) \propto \mathbb{P}(y_k \mid X_k)\, \mathbb{P}(X_k \mid \mathcal{Y}_{k-1}).$$

Bayesian Filtering Algorithm

The Bayesian filtering recursion can be summarized as follows:

1. **Initialization:** Specify a prior distribution over the initial state $\mathbb{P}(X_0)$.
2. **For** $k = 1, 2, \ldots, T$, repeat:
   (a) **Prediction:** Compute the predictive distribution

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

   (b) **Update:** Incorporate the new measurement y_k via Bayes' rule

   $$\mathbb{P}(X_k \mid \mathcal{Y}_k) \propto \mathbb{P}(y_k \mid X_k)\, \mathbb{P}(X_k \mid \mathcal{Y}_{k-1}).$$

   (c) **Normalization:** Normalize the distribution if required.
3. **Output:** The sequence of filtering distributions $\{\mathbb{P}(X_k \mid \mathcal{Y}_k)\}_{k=0}^{T}$.

Figure 1 provides a graphical interpretation of the Bayesian filtering recursion derived in the previous section.

At the initial time step, uncertainty about the system is represented by a prior distribution $\mathbb{P}(X_0)$. This prior encodes all available knowledge about the system state before any measurements are observed.

The horizontal arrows represent the *prediction* step. At each time step k, the filtering distribution from the previous time, $\mathbb{P}(X_{k-1} \mid \mathcal{Y}_{k-1})$, is propagated forward through the dynamics model to obtain the predictive distribution $\mathbb{P}(X_k \mid \mathcal{Y}_{k-1})$. This step accounts for the evolution of the system state and the accumulation of uncertainty due to process noise and model uncertainty.

The vertical arrows represent the *update* step. After the state evolves, a new measurement y_k is collected. This measurement provides information about the current state X_k and is incorporated

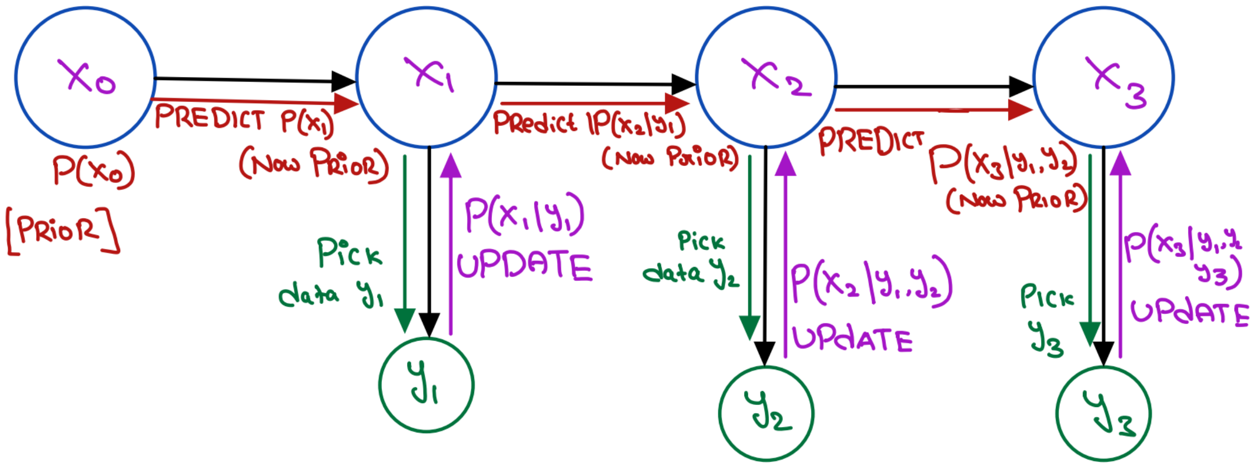

Figure 1: Prediction–update structure of Bayesian filtering. The posterior distribution at each time step is propagated forward through the dynamics model to form a predictive distribution, which is then updated using the newly acquired measurement to produce the filtering distribution.

through Bayes' rule to update the predictive distribution, yielding the filtering distribution $\mathbb{P}(X_k \mid \mathcal{Y}_k)$.

A key property of the filtering recursion is that all past measurements influence future estimates only through the current filtering distribution. That is, once $\mathbb{P}(X_k \mid \mathcal{Y}_k)$ has been computed, the entire measurement history $\mathcal{Y}_k$ can be summarized by this distribution.