DISCRETE-TIME KALMAN FILTERING

NAZIM FADLI

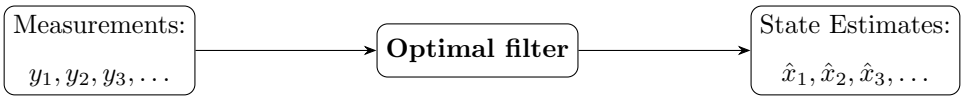
Contents

1. Introduction

These pages are part of a larger project I worked on during an internship. The detailed mathematical proofs for understanding the algorithms are not included here, but you can find them in the excellent references listed below, which I used to compile this material.

1.0.1. *The idea behind Optimal Filtering.*

Optimal filtering is a mathematical field focused on developing methods to achieve the most accurate estimate of the state of a dynamical system, which is indirectly observed through noisy measurements. Filtering can be viewed as a form of denoising, as it aims to extract the true underlying information from noisy observations to provide the most likely state of the system.



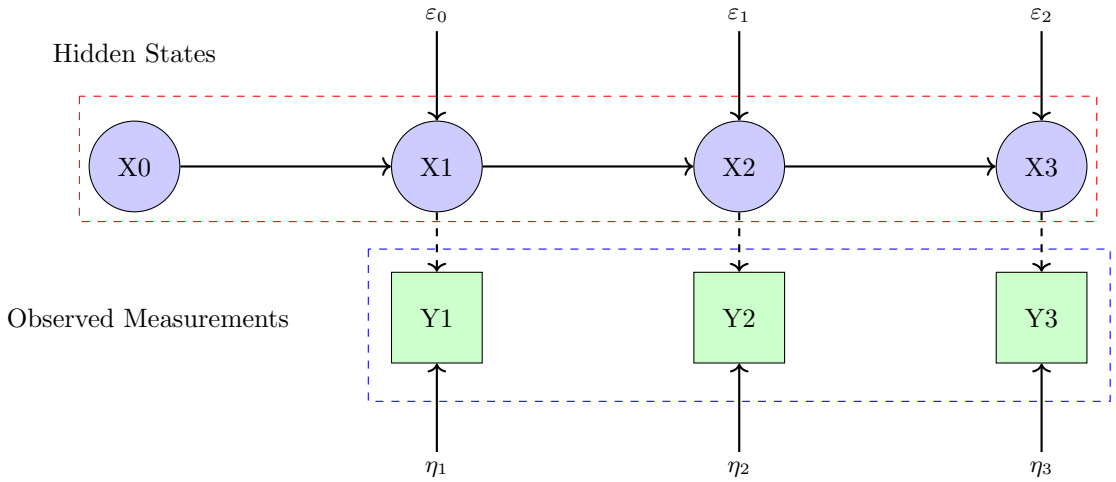
1.0.2. *The general framework.*

We are considering discrete-time systems of the form:

$$x_{k+1} = f(x_k, u_k, \varepsilon_k) \tag{1.1}$$

$$y_k = g(x_k, h_k, \eta_k) \tag{1.2}$$

- Equation (1.1) describes a **Dynamical System**[1] (it describes how the state evolves over time), where:
  (1) x_{k+1} is the state at the next time step (hidden random variable).
  (2) x_k is the current state, indirectly observed through measurements (hidden random variable).
  (3) f is a known function governing state transitions.
  (4) u_k is a known and deterministic system parameter.
  (5) ε_k represents process noise, modeling uncertainties in dynamics.
- Equation (1.2) represents a **Measurement Model** (it describes how the true state of the system is related to the observed measurements), where:
  (1) y_k is the measurement at time k, observed directly.
  (2) g is a known function that maps the system state to the observed measurements, which is known.
  (3) h_k is a known and deterministic measurement parameter.
  (4) η_k denotes measurement noise, accounting for uncertainties or errors due to sensor inaccuracies and variability.



Remarks 1.3.

- We often deal with Markovian dynamical systems.
- When a Markov model includes a measurement model that verifies the assumption of output independence [2], it forms a Hidden Markov Model (HMM).
- If our model is a Markov Model and we have a Measurement Model we call this combined structure a Hidden Markov model.
- Notice that even if the true system state x_k was known, the measurement y_k would not be a deterministic functions of the state but would have a distribution of possible values.
- The measurements space and the state space can be different from each other, for example, the state can be a vector of positions while the measurements can be a scalar that represents the total distance from a reference point.

[1]The random analogue of recursive sequences
[2]For $k \leq T, A \in \mathcal{A}, P(Y_k \in A \mid X_k) = P(Y_k \in A \mid X_1, \ldots, X_T, Y_1, \ldots, Y_T)$.

- The solution to the filtering problem depends significantly on the characteristics of the system, such as the distributions of the sequences of random variables $(\eta_k)_k$ and $(\varepsilon_k)_k$, as well as the properties of the functions g and f.
- We could also have an initial measurement at y_0.

For now, we will only assume that the sequence $(X_k)_{k\in\mathbb{N}}$ satisfies the Markov property. This means, given the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, if we denote by K the transition kernel of $(X_k)_{k\in\mathbb{N}}$: $\forall k \in \mathbb{N}, \forall A \in \mathcal{A}, \forall \omega \in \Omega$,

$$\mathbb{P}(X_{k+1} \in A \mid \mathcal{F}_k)(\omega) = \mathbb{P}(X_{k+1} \in A \mid X_k(\omega)) = K(X_k(\omega), A)$$

where $\mathcal{F}_k$ is the natural filtration of the process : $\mathcal{F}_k = \sigma(X_0, X_1, \ldots, X_k)$.

Additionally, we assume an **initial distribution**, which corresponds to the probability distribution $P(x_0)$ of the initial hidden state x_0. In some cases, the initial state can even be deterministic.

1.0.3. *Objective.*

We will focus on the Bayesian approach to filtering, where we aim to determine the following distributions:

- **Filtering distribution**: For all $k \in \mathbb{N}$ and $A \in \mathcal{A}$,

$$\mathbb{P}(X_k \in A \mid Y_1, Y_2, \ldots, Y_k).$$

- **Prediction distribution**: For all $k, n \in \mathbb{N}$ and $A \in \mathcal{A}$,

$$\mathbb{P}(X_{k+n} \in A \mid Y_1, Y_2, \ldots, Y_k).$$

- **Smoothing distribution**: For all $k, n \in \mathbb{N}$ such that $n > k$, and $A \in \mathcal{A}$,

$$\mathbb{P}(X_k \in A \mid Y_1, Y_2, \ldots, Y_n).$$

Given the filtering distribution, we usually use the mean of that distribution as an estimator for the state and can construct a confidence region.

2. Kalman Filter (KF)

When both the Dynamical System and the Measurement Model are linear Gausian we have a closed form solution to the Bayesian filtering problem [2].

Definition 2.1. The Kalman-Bucy Filter The Kalman Filter is an algorithm that iteratively improves estimates of a dynamical system's state using noisy measurements. It is a two-step algorithm that first predicts the state of a dynamic system and then corrects the prediction using noisy measurements.

2.0.1. *The Model.*

We consider a Linear Gaussian model for both measurement and dynamics for $k \geq 1$:

$$X_k = F_k X_{k-1} + u_k + \varepsilon_k \tag{2.2}$$

$$Y_k = H_k X_k + h_k + \eta_k \tag{2.3}$$

Where $F_k \in \mathbb{R}^{n\times n}$, $u_k \in \mathbb{R}^n$, $H_k \in \mathbb{R}^{m\times n}$, $h_k \in \mathbb{R}^m$ are all known parameters of the model.

Additionally, the model's assumptions are:

(1) The initial state X_0 follows a Gaussian distribution with known mean and variance.
(2) The process noise $(\varepsilon_k)_k$ is a sequence of independent Gaussian random variables with zero mean and known variance V_k^{ε}.
(3) The measurement noise $(\eta_k)_k$ is a sequence of independent Gaussian random variables with zero mean and known variance V_k^{η}.
(4) The measurement noise, the initial state, and the process noise are independent of one another.

We will define some notations:

$$\begin{aligned}
\hat{X}_{k|k} &= \mathbb{E}[X_k \mid \mathcal{F}_k] && : \text{the expected value of } X_k \text{ given the information } \mathcal{F}_k \\
\hat{X}_{k+1|k} &= \mathbb{E}[X_{k+1} \mid \mathcal{F}_k] && : \text{the expected value of } X_{k+1} \text{ given the information } \mathcal{F}_k \\
\mathcal{F}_k &= \sigma(Y_0, Y_1, \dots, Y_k) && : \text{the } \sigma\text{-algebra generated by the measurements up to time } k \\
P_{k|k} &= \mathbb{V}[X_k \mid \mathcal{F}_k] && : \text{the variance of } X_k \text{ given the information } \mathcal{F}_k \\
&= \mathbb{E}[(X_k - \hat{X}_{k|k})(X_k - \hat{X}_{k|k})^T]
\end{aligned}$$

The Kalman-Bucy Filter uses the principle that the conditional expectation $\mathbb{E}[X_k \mid \mathcal{F}_k]$ is the best L^2 approximation, minimizing the mean squared error for state estimates based on available information up to time k.

Theorem 2.4: ***The Kalman-Bucy Filter (KF) .*** *The Kalman-Bucy Filter is a two-step recursive algorithm. Suppose that the covariace matrix of measurement noise V_k^{η} is invertible for all $k \in \mathbb{N}$.*
Let $\hat{X}_{0|0} = \mu = \mathbb{E}[X_0]$ and $P_{0|0} = V = \mathbb{V}[X_0]$.

The prediction step is given by:

$$\hat{X}_{k+1|k} = F_{k+1}\hat{X}_{k|k} + u_{k+1}, \tag{2.5}$$

$$P_{k+1|k} = F_{k+1}P_{k|k}F_{k+1}^T + V_{k+1}^{\varepsilon} \quad \forall k > 0 \tag{2.6}$$

The correction step is given by:

$$\hat{X}_{k+1|k+1} = \hat{X}_{k+1|k} + K_{k+1}(Y_{k+1} - (H_{k+1}\hat{X}_{k+1|k} + h_{k+1})), \tag{2.7}$$

$$P_{k+1|k+1} = (I - K_{k+1}H_{k+1})\, P_{k+1|k} \tag{2.8}$$

where the Kalman gain K_{k+1} is defined as:

$$K_{k+1} = P_{k+1|k}H_{k+1}^T \left(H_{k+1}P_{k+1|k}H_{k+1}^T + V_{k+1}^{\eta}\right)^{-1} \tag{2.9}$$

Thus, the state X_k given $\mathcal{F}_k$ is Gaussian distributed as $\mathcal{N}(\hat{X}_{k|k}, P_{k|k})$, and the distribution of X_{k+1} given $\mathcal{F}_k$ is Gaussian distributed as $\mathcal{N}(\hat{X}_{k+1|k}, P_{k+1|k})$.

2.1. **Implementation.** The Kalman filter algorithm can be implemented as follows:

Algorithm 1 Kalman Filter Algorithm

Initialization:
Initial state estimate: $\hat{X}_{0|0}$
Initial covariance estimate: $P_{0|0}$
for $k = 0, 1, 2, \ldots$ **do**
  Prediction Step:
  Predicted state estimate:
$$\hat{X}_{k+1|k} = F_k \hat{X}_{k|k} + u_k$$
  Predicted covariance estimate:
$$P_{k+1|k} = F_k P_{k|k} F_k^T + V_k^{\varepsilon}$$
  Update Step:
  Kalman gain:
$$K_{k+1} = P_{k+1|k} H_k^T \left(H_k P_{k+1|k} H_k^T + V_k^{\eta} \right)^{-1}$$
  Innovation:
$$I_{k+1} = Y_{k+1} - H_k \hat{X}_{k+1|k} - h_k$$
  Updated state estimate:
$$\hat{X}_{k+1|k+1} = \hat{X}_{k+1|k} + K_{k+1} I_{k+1}$$
  Updated covariance estimate:
$$P_{k+1|k+1} = (I - K_{k+1} H_k) P_{k+1|k}$$
end for

3. Kalman Smoother (KS)

We will now discuss the solution to the smoothing problem for the same model described in Section 2.0.1. The key distinction between filters and smoothers lies in the timing of the measurements they utilize for estimating states. A Bayesian filter computes its estimates based solely on measurements obtained up to and including the current time step k. In contrast, a Bayesian smoother refines its estimates by incorporating future measurements.

Since we have already presented the Kalman Filter, we will now formulate the Kalman Smoother as a forward-backward smoother.This approach assumes that the Kalman Filter has already been applied to our data (alternative formulations exist that do not require the use of the Kalman Filter) and that for all $k \in \mathbb{N}$, both noise covariance matrices V_k^{η} and V_k^{ε} are invertible. Although it is possible to formalize the Kalman Smoother without assuming the invertibility of the noise process covariance (Fraser–Potter formulation, see [1])[3], we will proceed with this assumption for the sake of simplicity.

We will define for for $n \in \mathbb{N}, k \in \mathbb{N}$:

$M_0^n = (Y_0, \ldots, Y_n)$: The random vector representing the history of measurements

$E_{k+1}^n = (\varepsilon_{k+1}, \ldots, \varepsilon_n, \eta_k, \eta_{k+1}, \ldots, \eta_n)$: The random vector of noise history from step k to step n

$\hat{X}_{k|n} = \mathbb{E}[X_k \mid \mathcal{F}_n]$: the expected value of X_k given the information

[3]With this formulation, we can reduce data storage requirements.

Theorem 3.1: ***The Kalman Smoother (KS)***[3] **.** *Also known as the Rauch–Tung–Striebel smoother, the Kalman Smoother is a two-step algorithm. Assuming that the Kalman Filter has already been applied to the data, the Kalman Smoother involves a backward recursion step to refine the state estimates.*

Suppose that $k \in \mathbb{N}$ and $n > k$, and that the Matrices V_k^η and V_k^ε are invertible for all $k \in \mathbb{N}$.

Let our initial Kalman smoother estimates $\hat{X}_{n|n}$ and $P_{n|n}$ be set to the Kalman filter estimates at time n[4]*.*

The prediction step is given by [5]*:*

$$\hat{X}_{k+1|k} = F_k \hat{X}_{k|k} + u_k, \tag{3.2}$$

$$P_{k+1|k} = F_k P_{k|k} F_k^T + V_k^\varepsilon \quad \forall k \geq 0 \tag{3.3}$$

The backward recursion step is given by:

$$\hat{X}_{k|n} = \hat{X}_{k|k} + S_k(\hat{X}_{k+1|n} - \hat{X}_{k+1|k}), \tag{3.4}$$

$$P_{k|n} = P_{k|k} + S_k(P_{k+1|n} - P_{k+1|k})S_k^T \tag{3.5}$$

where the smoothing gain S_k is defined as:

$$S_k = P_{k|k} F_k^T P_{k+1|k}^{-1} \tag{3.6}$$

Thus, the state X_k given $\mathcal{F}_n$ is Gaussian distributed as $\mathcal{N}(\hat{X}_{k|n}, P_{k|n})$.

Remarks 3.7.

- Note that we are using the predicted covariance $P_{k+1|k}$ and the predicted state estimate $\hat{X}_{k+1|k}$, which are computed by the Kalman filter.
- We can show that $P_{k|k} - P_{k|n}$ is positive semidefinite.
- The recursion can be used to compute the smoothing distributions for all time steps by starting from the last step n and proceeding backward to the initial step $k = 0$.
- Note that at the final step n, the smoothing estimates are the same as the Kalman filter estimates : $P_{n|n}$ and $\hat{X}_{n|n}$.

[4]The initialization step is at time T

[5]The Kalman filter prediction equations

3.1. **Implementation.** The Kalman smoother algorithm can be implemented as follows:

Algorithm 2 Kalman Smoother Algorithm

Initialization:
Initial state estimate: $\hat{X}_{0|0}$
Initial covariance estimate: $P_{0|0}$
for $k = 0, 1, 2, \ldots, T-1$ **do**
    Prediction Step (using Kalman filter) 2.1
    Update Step (using Kalman filter)
end for
Smoothing Step:
Initial smoothed estimates: $\hat{X}_{T|T}$, $P_{T|T}$
for $k = T-1, T-2, \ldots, 0$ **do**
    Smoothing gain:

$$J_k = P_{k|k} F_k^T \left(P_{k+1|k}\right)^{-1}$$

    Smoothed state estimate:

$$\hat{X}_{k|T} = \hat{X}_{k|k} + J_k \left(\hat{X}_{k+1|T} - \hat{X}_{k+1|k}\right)$$

    Smoothed covariance estimate:

$$P_{k|T} = P_{k|k} + J_k \left(P_{k+1|T} - P_{k+1|k}\right) J_k^T$$

end for

3.2. **Pros and Cons of the Kalman Filter.**

+ Provides an analytic and exact solution to the filtering problem, making it easy to implement
+ Offers the best approximation according to the mean squared error (optimal solution)
- Only works with Linear Gaussian model for both measurement and dynamics

4. Extension

There are a lot of extensions to the Kalman Filter for non-linear or Gaussian models (Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF)...).

References

[1] François Le Gland. Introduction au filtrage en temps discret: Filtrage de kalman et modèles de markov cachés. Lecture notes for Master EEEA — Spécialité SISEA (Signal, Image, Systèmes Embarqués, Automatique). Available at `http://www.irisa.fr/aspi/legland/rennes-1/`.

[2] Andrew H. Jazwinski, editor. *Stochastic Processes and Filtering Theory*, volume 64. Academic Press, Seabrook, Maryland, 1970.

[3] Simo Särkkä and Lennart Svensson. *Bayesian Filtering and Smoothing*. Cambridge University Press, second edition, 2023.