

Notes on the Kalman Filter

Sehwan Jhang

September 2024

1 Preliminaries

The Kalman filter is an algorithm that recursively predicts and updates the state of a time-varying system based on predictions from the system model and noisy observations. This section provides a brief overview of the foundational knowledge before delving into the Kalman filter itself.

1.1 Estimation Theory

Estimation theory encompasses various methods for predicting the parameters of a model based on observed data. It is widely used in fields such as data analysis, machine learning, finance, and robotics, serving as an essential tool for making accurate predictions in the presence of uncertainty.

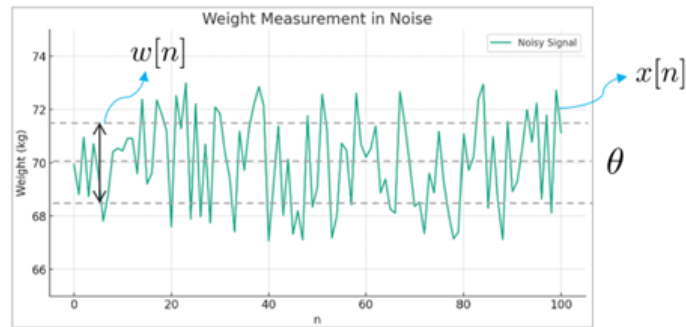


Figure 1: Example of data used in estimation theory.

For example, suppose we are given data measured over a year for a person whose average weight is 70 kg, as shown in the figure above. The green

data points forming the graph are $x[n]$, and the central value of these data points is the parameter θ that we aim to estimate. The degree of spread (i.e., variance) around θ can be mathematically represented by $w[n]$. The problem of estimating the parameter θ based on the observed data $x[n]$ can be expressed by the following simple mathematical model:

$$x[n] = \theta + w[n] = 0, 1, \dots, N - 1 \quad (1)$$

- $x[n]$: observed data,
- θ : parameter to be estimated,
- $w[n]$: random noise.

1.2 Bayesian Philosophy

Estimation theory can be broadly divided into two perspectives: the frequentist and the Bayesian, depending on how the parameter θ to be estimated is viewed.

- Frequentist : A frequentist perspective views the parameter θ as an unknown deterministic parameter.
- Bayesian : A Bayesian perspective treats the parameter θ as a random variable with a prior probability distribution.

$$\begin{aligned} \text{Frequentist: } \underbrace{x[n]}_{\text{r.v.}} &= \underbrace{\theta}_{\text{deterministic}} + \underbrace{w[n]}_{\text{r.v.}} \\ \text{Bayesian: } \underbrace{x[n]}_{\text{r.v.}} &= \underbrace{\theta}_{\text{r.v.}} + \underbrace{w[n]}_{\text{r.v.}} \end{aligned} \quad (2)$$

The Bayesian philosophy starts from the premise that if we have prior information about the parameter θ , it can be used to make a better estimate. For this, a prior pdf of θ must be provided or computable. In the Bayesian framework, since the parameter θ can also be modeled as a random variable, the following Bayesian rule holds:

$$p(\theta | x) = \frac{p(\theta)p(x | \theta)}{p(x)} \quad (3)$$

- $p(\theta | x)$: posterior conditional probability distribution of the parameter θ given the observed data x

- $p(x \mid \theta)$: conditional probability distribution of x given θ , or likelihood
- $p(x)$: probability distribution of x
- $p(\theta)$: prior probability distribution of θ

1.3 Estimation problem

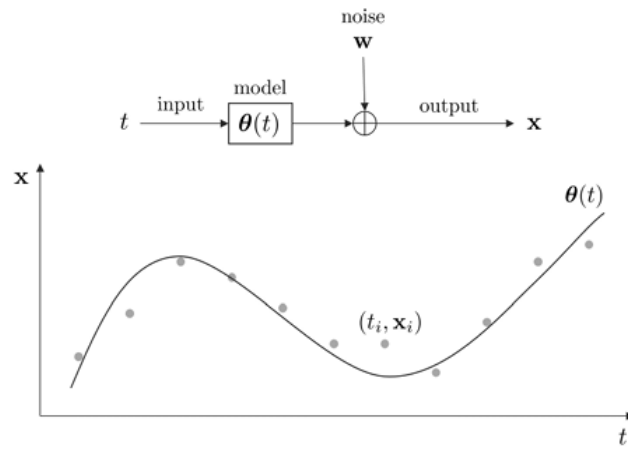


Figure 2: Example of data used in estimation theory.

The goal of the estimation problem is to find the optimal model $\theta(t)$ using the given data (t_i, x_i) . The type of estimation problem varies depending on which point in time is being estimated.

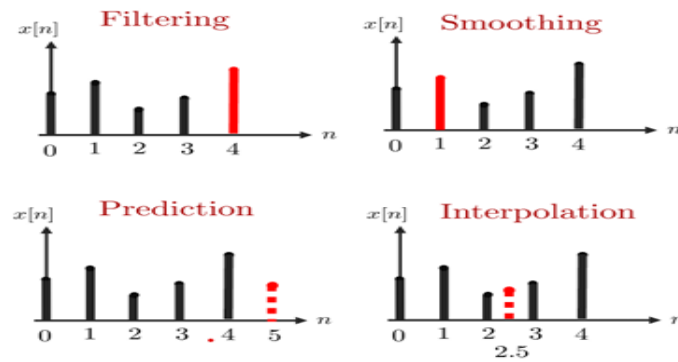


Figure 3: Example of data used in estimation theory.

1. **Filtering:** Filtering refers to the problem of estimating $\theta = x[N - 1]$ given the observed data $\{x[0], x[1], \dots, x[N - 1]\}$. By estimating the optimal parameter, we aim to filter out the noise from the signal. Note that in filtering, the parameter depends only on current and past data.
2. **Smoothing:** Smoothing refers to the problem of estimating an intermediate $\theta = s[n]$ given the observed data $\{x[0], x[1], \dots, x[N - 1]\}$. For instance, to estimate $s[1]$, all observed data are used. Naturally, smoothing cannot be performed until all data are observed.
3. **Prediction:** Prediction refers to the problem of estimating $\theta = x[N - 1 + l]$ given the observed data $\{x[0], x[1], \dots, x[N - 1]\}$. Here, l is an arbitrary positive number.