

Understanding Extended Kalman Filter – Part I: One Dimensional Kalman Filter

Shoudong Huang

ARC Centre of Excellence for Autonomous Systems (CAS)
Faculty of Engineering and Information Technology,
University of Technology Sydney
Email: sdhuang@eng.uts.edu.au

April 23, 2010

Abstract

Kalman Filter (KF) and extended Kalman Filter (EKF) are basic tools for solving many estimation problems. They have found plenty of applications including target tracking and mobile robot localization and mapping. In this note, I tried to explain the KF and EKF formulas without using too much knowledge of probability theory. In Part I (this note), one dimensional KF formula is derived. In Part II, multi dimensional KF formula is proven and in Part III, EKF formula is proven.

Contents

1	Introduction	2
2	Kalman Filter formula	2
3	1D Gaussian distribution and its information	3
4	Important properties of 1D Gaussian distributions	3
5	1D Kalman Filter prediction	4
5.1	The simplest process model	4
5.2	General process model	5
6	1D Kalman Filter update	5
6.1	The simplest observation model	6
6.2	The general observation model	8

1 Introduction

The structure of this note is the following. The Kalman Filter formula is first reviewed in Section 2. One dimensional Gaussian distribution is reviewed in Section 3 and some of its properties are listed in Section 4. The derivation of 1D Kalman filter prediction formulas are explained in Section 5 and the 1D Kalman filter update formulas are derived in Section 6.

I am assuming the readers know a bit on KF and EKF formulas but does not quite understand how to derive them. If you know nothing about KF and EKF, you can still try to read it. If you are an expert on KF and EKF, reading this note may not benefit yourself but will benefit me (you can provide good comments/suggestions to me).

I hope this note can help you understand KF and EKF a bit more. If you find it is too difficult to follow after spending 30 minutes reading it, then throw it away (this means the note is not suitable for you). However, it will be great if you can let me know why it is difficult to follow.

2 Kalman Filter formula

For a linear system, the process model (from time k to time $k + 1$) is described as

$$\mathbf{x}_{k+1} = F\mathbf{x}_k + G\mathbf{u}_k + \mathbf{w}_k, \quad (1)$$

where $\mathbf{x}_k, \mathbf{x}_{k+1}$ are the system state (vector) at time $k, k + 1$, F is the system transition matrix, G is the gain of control \mathbf{u}_k , and \mathbf{w}_k is the zero-mean Gaussian process noise $\mathbf{w}_k \sim N(0, Q)$.

For state estimation problem, the true system state is not available and needs to be estimated. The initial state x_0 is assumed to follow a known Gaussian distribution $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, P_0)$. The objective is to estimate the state at each time step by the process model and the observations.

The observation model at time $k + 1$ is given by

$$\mathbf{z}_{k+1} = H\mathbf{x}_{k+1} + \mathbf{v}_{k+1}. \quad (2)$$

where H is the observation matrix and \mathbf{v}_{k+1} is the zero-mean Gaussian observation noise $\mathbf{v}_{k+1} \sim N(0, R)$.

Suppose the knowledge on \mathbf{x}_k at time k (after the observation at time k) is

$$\mathbf{x}_k \sim N(\hat{\mathbf{x}}_k, P_k), \quad (3)$$

then \mathbf{x}_{k+1} at time $k + 1$ (after the observation at time k) follows

$$\mathbf{x}_{k+1} \sim N(\hat{\mathbf{x}}_{k+1}, P_{k+1}) \quad (4)$$

where the state estimate $\hat{\mathbf{x}}_{k+1}$ and its corresponding covariance matrix P_{k+1} can be computed by the following **Kalman Filter** formula.

Predict using process model:

$$\bar{\mathbf{x}}_{k+1} = F\hat{\mathbf{x}}_k + G\mathbf{u}_k \quad (5)$$

$$\bar{P}_{k+1} = FP_kF^T + Q \quad (6)$$

where $\bar{\mathbf{x}}_{k+1}$ is the state estimate at time $k + 1$ before using the observation information at time $k + 1$, and \bar{P}_{k+1} is its corresponding covariance matrix.

Update using observation:

$$\hat{\mathbf{x}}_{k+1} = \bar{\mathbf{x}}_{k+1} + K(\mathbf{z}_{k+1} - H\bar{\mathbf{x}}_{k+1}) \quad (7)$$

$$P_{k+1} = \bar{P}_{k+1} - KSK^T, \quad (8)$$

where the innovation covariance S (here $\mathbf{z}_{k+1} - H\bar{\mathbf{x}}_{k+1}$ is called innovation) and the Kalman gain K are given by

$$S = H\bar{P}_{k+1}H^T + R \quad (9)$$

$$K = \bar{P}_{k+1}H^T S^{-1}. \quad (10)$$

Remark 2.1 If you just want to apply Kalman Filter, then the above formulas are enough as long as you have got your process model and observation model. If you would like to know where these formulas come from, then proceed to the following sections.

3 1D Gaussian distribution and its information

If a random variable x follows a Gaussian distribution, it is denoted as

$$x \sim N(m, \sigma^2) \quad (11)$$

where m is the mean and σ^2 is the variance.

The meaning is — x is likely to be around the mean m , the level of uncertainty depends on the variance σ^2 . The larger the variance, the larger the uncertainty. See Figure 1.

(Fisher) information of a Gaussian distribution $N(m, \sigma^2)$ is the inverse of the variance,

$$I = \frac{1}{\sigma^2}. \quad (12)$$

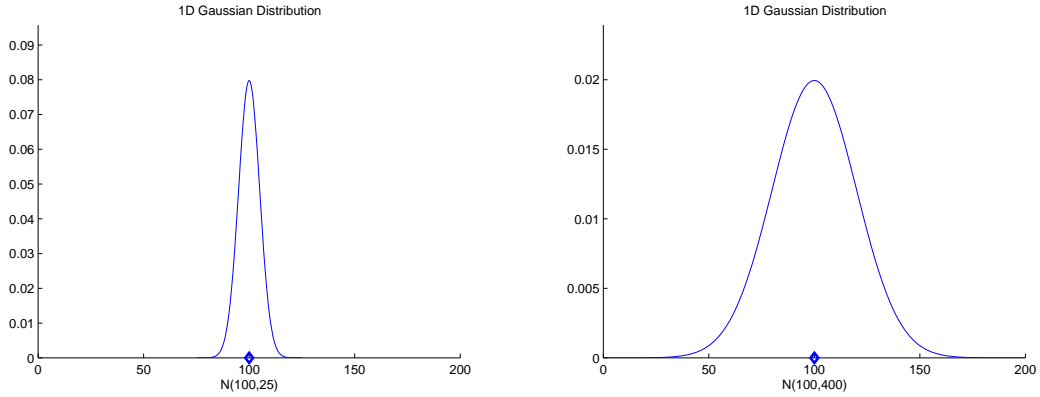
The larger the uncertainty, the smaller the information. Also see Figure 1.

4 Important properties of 1D Gaussian distributions

Below are a few properties of 1D Gaussian distributions which are useful in deriving the Kalman Filter formulas.

- For any constant a ,

$$x \sim N(m, \sigma^2) \Rightarrow ax \sim N(am, a^2\sigma^2). \quad (13)$$



(a) One dimensional Gaussian distribution with smaller variance $\sigma = 5$ (information = $1/25$). (b) One dimensional Gaussian distribution with larger variance $\sigma = 20$ (information = $1/400$).

Figure 1: One dimensional Gaussian distributions

- For any constant u ,

$$x \sim N(m, \sigma^2) \Rightarrow x + u \sim N(m + u, \sigma^2). \quad (14)$$

- For two independent random variable x and y (the value of x contains no information about the value of y and vice versa),

$$x \sim N(m_x, \sigma_x^2), y \sim N(m_y, \sigma_y^2) \Rightarrow x + y \sim N(m_x + m_y, \sigma_x^2 + \sigma_y^2). \quad (15)$$

Remark 4.1 These properties agree with our intuition. The rigorous proof of these properties can be found in many tutorials or books on probability. For example, the following website:

http://en.wikipedia.org/wiki/Gaussian_distribution

5 1D Kalman Filter prediction

This section shows that the KF prediction formula can be obtained easily from the properties of 1D Gaussian distributions listed in Section 4.

5.1 The simplest process model

Suppose the process model is

$$x_{k+1} = x_k + u_k + w_k \quad (16)$$

where u_k is the control (a constant from time k to time $k + 1$) and w_k is the zero-mean Gaussian process noise with variance σ_u^2 . That is $w_k \sim N(0, \sigma_u^2)$. It is also assumed that w_k is independent of x_k .

At time k , the estimate of x_k follows a Gaussian distribution $x_k \sim N(\hat{x}_k, \sigma_k^2)$ (see equation (3)), thus by property (14),

$$x_k + u_k \sim N(\hat{x}_k + u_k, \sigma_k^2). \quad (17)$$

Further by property (15),

$$x_{k+1} = (x_k + u_k) + w_k \sim N(\hat{x}_k + u_k, \sigma_k^2 + \sigma_u^2). \quad (18)$$

Thus if we denote the estimate of x_{k+1} (after the process but before the observation) as

$$x_{k+1} \sim N(\bar{x}_{k+1}, \bar{\sigma}_{k+1}^2), \quad (19)$$

then **the prediction formula** is

$$\begin{aligned} \bar{x}_{k+1} &= \hat{x}_k + u_k, \\ \bar{\sigma}_{k+1}^2 &= \sigma_k^2 + \sigma_u^2. \end{aligned} \quad (20)$$

This is the formula (5) and (6) (when all variables are scalars).

5.2 General process model

Similarly, if the process model is

$$x_{k+1} = ax_k + bu_k + w_k \quad (21)$$

where a, b are constant and $w_k \sim N(0, \sigma_u^2)$.

Since $x_k \sim N(\hat{x}_k, \sigma_k^2)$, by property (13),

$$ax_k \sim N(a\hat{x}_k, a^2\sigma_k^2). \quad (22)$$

By property (14) (here bu_k is a constant),

$$ax_k + bu_k \sim N(a\hat{x}_k + bu_k, a^2\sigma_k^2). \quad (23)$$

Further by property (15),

$$x_{k+1} = (ax_k + bu_k) + w_k \sim N(a\hat{x}_k + bu_k, a^2\sigma_k^2 + \sigma_u^2). \quad (24)$$

Thus if we denote the estimate of x_{k+1} (after the process but before the observation) as

$$x_{k+1} \sim N(\bar{x}_{k+1}, \bar{\sigma}_{k+1}^2), \quad (25)$$

then **the prediction formula** is

$$\begin{aligned} \bar{x}_{k+1} &= a\hat{x}_k + bu_k, \\ \bar{\sigma}_{k+1}^2 &= a^2\sigma_k^2 + \sigma_u^2. \end{aligned} \quad (26)$$

This is the formula (5) and (6) (when all variables are scalars).

6 1D Kalman Filter update

This section shows that the KF update formula can be obtained easily by adding the information from observation to the prior information.

6.1 The simplest observation model

Suppose the observation model is

$$z_{k+1} = x_{k+1} + v_{k+1}, \quad (27)$$

where z_{k+1} is the observation value at time $k + 1$ (constant) and v_{k+1} is the zero-mean Gaussian observation noise with variance σ_z^2 . That is $v_{k+1} \sim N(0, \sigma_z^2)$. It is also assumed that v_{k+1} is independent of x_{k+1} . By property (13) (choosing $a = -1$),

$$-v_{k+1} \sim N(0, \sigma_z^2). \quad (28)$$

By the observation model (27),

$$x_{k+1} = -v_{k+1} + z_{k+1}. \quad (29)$$

Thus by property (14) (add a constant z_{k+1}),

$$x_{k+1} \sim N(z_{k+1}, \sigma_z^2). \quad (30)$$

The prior information about x_{k+1} is given by (25) (after the prediction but before the update). So we have two pieces of information about x_{k+1} — one from observation (30) and one from prior (25).

According to the definition of information contained in a Gaussian distribution (see (12) in Section 3), the information (about x_{k+1}) contained in (25) is

$$I_{prior} = \frac{1}{\bar{\sigma}_{k+1}^2}, \quad (31)$$

while the information (about x_{k+1}) contained in (30) is

$$I_{obs} = \frac{1}{\sigma_z^2}. \quad (32)$$

The total information (about x_{k+1}) after the observation should be the sum of the two, namely,

$$I_{total} = I_{prior} + I_{obs} = \frac{1}{\bar{\sigma}_{k+1}^2} + \frac{1}{\sigma_z^2}. \quad (33)$$

The new mean value is the weighted sum of the mean values of the two Gaussian distributions (30) and (25). The weights are decided by the proportion of information contained in each of the Gaussian distributions (as compared with the total information). That is,

$$\hat{x}_{k+1} = \frac{I_{prior}}{I_{total}} \bar{x}_{k+1} + \frac{I_{obs}}{I_{total}} z_{k+1} = \frac{\sigma_z^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2} \bar{x}_{k+1} + \frac{\bar{\sigma}_{k+1}^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2} z_{k+1}. \quad (34)$$

(Note that the sum of the two weights is one, i.e. $\frac{I_{prior}}{I_{total}} + \frac{I_{obs}}{I_{total}} = 1$. Moreover, when the observation contains less information, the new mean is closer to the prior mean \bar{x}_{k+1} ; when the prior contains less information, the new mean is closer to the observation z_{k+1} .)

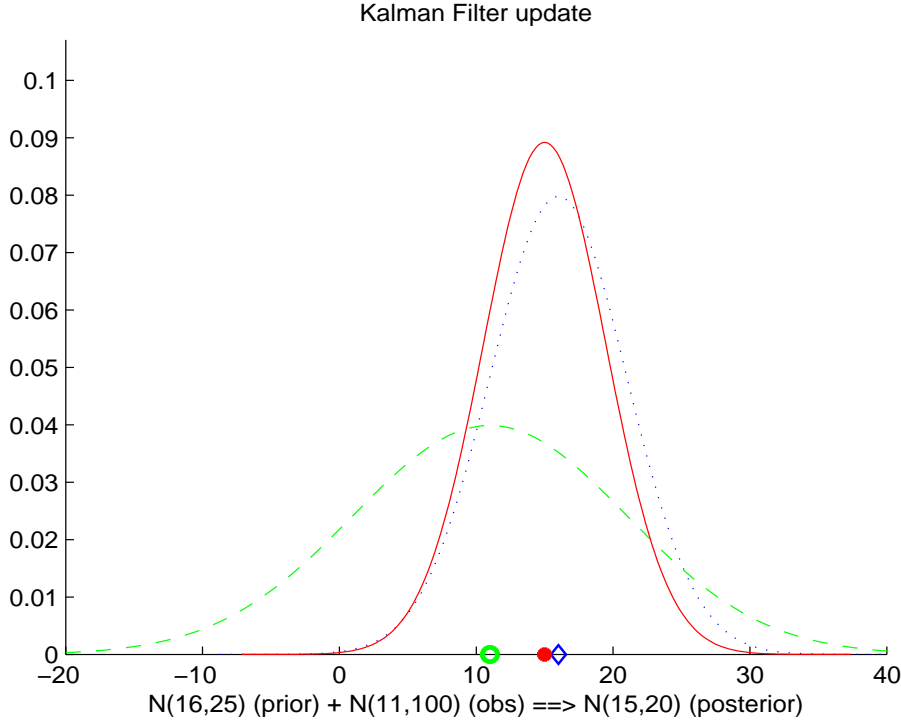


Figure 2: One dimensional KF update (dot line is the prior, dashed line is the observation, solid line is the posterior). $I_{prior} = 1/25$, $I_{obs} = 1/100$, $I_{total} = I_{prior} + I_{obs} = 1/20$; posterior variance = $\frac{1}{I_{total}} = 20$; posterior mean = $\frac{I_{prior}}{I_{total}} \times 16 + \frac{I_{obs}}{I_{total}} \times 11 = 15$.

The variance can be obtained by (see (12) in Section 3)

$$\sigma_{k+1}^2 = \frac{1}{I_{total}} = \frac{1}{\frac{1}{\bar{\sigma}_{k+1}^2} + \frac{1}{\sigma_z^2}} = \frac{\bar{\sigma}_{k+1}^2 \sigma_z^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2}. \quad (35)$$

So, the final estimate on x_{k+1} (after the prediction and update) is

$$x_{k+1} \sim N(\hat{x}_{k+1}, \hat{\sigma}_{k+1}^2) \quad (36)$$

where \hat{x}_{k+1} and $\hat{\sigma}_{k+1}^2$ are given in the above equations (34) and (35).

Figure 2 illustrates the update step of Kalman Filter.

The update equations (34) and (35) can also be expressed as

$$\begin{aligned} \hat{x}_{k+1} &= \bar{x}_{k+1} + \frac{\bar{\sigma}_{k+1}^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2} (z_{k+1} - \bar{x}_{k+1}) \\ \sigma_{k+1}^2 &= \bar{\sigma}_{k+1}^2 \left(1 - \frac{\bar{\sigma}_{k+1}^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2}\right) \\ &= \bar{\sigma}_{k+1}^2 - \bar{\sigma}_{k+1}^2 \frac{\bar{\sigma}_{k+1}^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2} \\ &= \bar{\sigma}_{k+1}^2 - \frac{\bar{\sigma}_{k+1}^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2} (\bar{\sigma}_{k+1}^2 + \sigma_z^2) \frac{\bar{\sigma}_{k+1}^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2}. \end{aligned} \quad (37)$$

This is the formula (7) and (8) (when all variables are scalars), where $z_{k+1} - \bar{x}_{k+1}$ is the innovation, $\bar{\sigma}_{k+1}^2 + \sigma_z^2$ is the innovation variance S , and $\frac{\bar{\sigma}_{k+1}^2}{\bar{\sigma}_{k+1}^2 + \sigma_z^2}$ is the Kalman gain K .

6.2 The general observation model

Suppose the observation model is

$$z_{k+1} = cx_{k+1} + v_{k+1}, \quad (38)$$

where c is a constant and $v_{k+1} \sim N(0, \sigma_z^2)$.

By the observation model (38),

$$x_{k+1} = -c^{-1}v_{k+1} + c^{-1}z_{k+1}. \quad (39)$$

By property (13) (choosing $a = -c^{-1}$),

$$-c^{-1}v_{k+1} \sim N(0, c^{-2}\sigma_z^2). \quad (40)$$

Thus by property (14) (add a constant $c^{-1}z_{k+1}$),

$$x_{k+1} \sim N(c^{-1}z_{k+1}, c^{-2}\sigma_z^2). \quad (41)$$

The prior information about x_{k+1} is given by (25) (after the prediction but before the update). So we have two pieces information about x_{k+1} — the information from observation (41) and the prior information (25).

According to the definition of information contained in a Gaussian distribution (see (12) in Section 3), the information (about x_{k+1}) contained in the prior (25) is

$$I_{prior} = \frac{1}{\bar{\sigma}_{k+1}^2}, \quad (42)$$

while the information (about x_{k+1}) contained in (41) is

$$I_{obs} = \frac{1}{c^{-2}\sigma_z^2}. \quad (43)$$

The total information (about x_{k+1}) after the observation should be the sum of the two, namely,

$$I_{total} = I_{prior} + I_{obs} = \frac{1}{\bar{\sigma}_{k+1}^2} + \frac{1}{c^{-2}\sigma_z^2}. \quad (44)$$

The new mean value is the weighted sum of the mean values of the two Gaussian distributions (41) and (25). The weights are decided by the proportion of information contained in each of the Gaussian distributions (as compared with the total information). That is,

$$\hat{x}_{k+1} = \frac{I_{prior}}{I_{total}}\bar{x}_{k+1} + \frac{I_{obs}}{I_{total}}(c^{-1}z_{k+1}) = \frac{\sigma_z^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2}\bar{x}_{k+1} + \frac{c\bar{\sigma}_{k+1}^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2}z_{k+1}. \quad (45)$$

The final variance can be obtained by (see (12) in Section 3)

$$\sigma_{k+1}^2 = \frac{1}{I_{total}} = \frac{1}{\frac{1}{\bar{\sigma}_{k+1}^2} + \frac{1}{c^{-2}\sigma_z^2}} = \frac{\bar{\sigma}_{k+1}^2\sigma_z^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2}. \quad (46)$$

So, the final estimate on x_{k+1} (after the prediction and update) is

$$x_{k+1} \sim N(\hat{x}_{k+1}, \hat{\sigma}_{k+1}^2) \quad (47)$$

where \hat{x}_{k+1} and $\hat{\sigma}_{k+1}^2$ are given in the above update equations.

The update equations (45) and (46) can also be expressed as

$$\begin{aligned} \hat{x}_{k+1} &= \bar{x}_{k+1} + \frac{c\bar{\sigma}_{k+1}^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2} (z_{k+1} - c\bar{x}_{k+1}) \\ \sigma_{k+1}^2 &= \bar{\sigma}_{k+1}^2 \left(1 - \frac{c^2\bar{\sigma}_{k+1}^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2}\right) \\ &= \bar{\sigma}_{k+1}^2 - \bar{\sigma}_{k+1}^2 \frac{c^2\bar{\sigma}_{k+1}^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2} \\ &= \bar{\sigma}_{k+1}^2 - \frac{c\bar{\sigma}_{k+1}^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2} (c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2) \frac{c\bar{\sigma}_{k+1}^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2}. \end{aligned} \quad (48)$$

This is the formula (7) and (8) (when all variables are scalars), where $z_{k+1} - c\bar{x}_{k+1}$ is the innovation, $c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2$ is the innovation variance S , and $\frac{c\bar{\sigma}_{k+1}^2}{c^2\bar{\sigma}_{k+1}^2 + \sigma_z^2}$ is the Kalman gain K .

Remark 6.1 If you have read through all the sections and they do make sense to you, then congratulations to you. If you feel this is not enough and you also want to understand more on multi-dimensional KF and even EKF, then please read Part II (and Part III).