# Understanding Extended Kalman Filter – Part II: Multi-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

March 18, 2009

**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, one dimensional KF formula is derived (you should have read it already). In Part II (this note), multi dimensional KF formula is proven and in Part III, EKF formula is proven.

## Contents

# 1 Introduction

I assume the readers have already read the Part I of this notes (One dimensional KF).

The structure of this note is the following. The Kalman Filter formula is reviewed again in Section 2. The multi-dimensional Gaussian distribution is reviewed in Section 3 and some of its properties are listed in Section 4. The derivation of multi-dimensional Kalman filter prediction formulas are explained in Section 5 and the multi-dimensional Kalman filter update formulas are derived in Section 6. Section 7 provides the matrix inversion lemma which is used in the derivation of the formulas.

# 2 Kalman Filter formula

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

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

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

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

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 $k$ is

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

then $\mathbf{x}_{k+1}$ at time $k+1$ follows

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

where $\hat{\mathbf{x}}_{k+1}, P_{k+1}$ can be computed by the following **Kalman Filter** formula.

**Predict** using process model:

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

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

**Update** using observation:

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

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

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

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

$$K = \bar{P}_{k+1}H^TS^{-1}. \tag{10}$$

**Remark 2.1** If you just want to apply Kalman Filter, then the above formulae 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 Gaussian distribution and information matrix

If a random vector $\mathbf{x}$ follows a Gaussian distribution, it is denoted as

$$\mathbf{x} \sim N(\mathbf{m}, P) \tag{11}$$

where $\mathbf{m}$ is the mean and $P$ is the covariance matrix.

The meaning is — $\mathbf{x}$ is likely to be around the mean $\mathbf{m}$, the level and shape of the uncertainty depends on the covariance matrix $P$. Here $P$ is a positive definite matrix.

The associated **Fisher information matrix** of Gaussian distribution $N(\mathbf{m}, P)$ is

$$I = P^{-1}. \tag{12}$$

Note that this information is a matrix instead of a scalar. So a quantity measure (for example, determinant, trace, largest eigenvalue) is necessary to compare two information matrices.

# 4 Important properties of Gaussian distributions

Below are a few properties of Gaussian distributions which are useful in deriving Kalman Filter.

- For any constant matrix $F$,

$$\mathbf{x} \sim N(\mathbf{m}, P) \Rightarrow F\mathbf{x} \sim N(F\mathbf{m}, FPF^T). \tag{13}$$

- For any constant vector $\mathbf{u}$,

$$\mathbf{x} \sim N(\mathbf{m}, P) \Rightarrow \mathbf{x} + \mathbf{u} \sim N(\mathbf{m} + \mathbf{u}, P). \tag{14}$$

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

$$\mathbf{x} \sim N(\mathbf{m}_x, P_x), \mathbf{y} \sim N(\mathbf{m}_y, P_y) \Rightarrow \mathbf{x} + \mathbf{y} \sim N(\mathbf{m}_x + \mathbf{m}_y, P_x + P_y). \tag{15}$$

**Remark 4.1** These properties are extensions of the properties for one dimensional Gaussian distributions. 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 Kalman Filter prediction

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

Suppose the process model is

$$\mathbf{x}_{k+1} = F\mathbf{x}_k + G\mathbf{u}_k + \mathbf{w}_k \tag{16}$$

where $\mathbf{u}_k$ is the control vector (a constant vector from time $k$ to time $k+1$) and $w_k$ is the zero-mean Gaussian process noise vector with covariance matrix $Q$. That is $\mathbf{w}_k \sim N(0, Q)$. It is also assumed that $\mathbf{w}_k$ is independent with $\mathbf{x}_k$.

At time $k$, the estimate of $\mathbf{x}_k$ is a Gaussian distribution $\mathbf{x}_k \sim N(\hat{\mathbf{x}}_k, P_k)$ (see equation (3)), thus by property (13),

$$F\mathbf{x}_k \sim N(F\hat{\mathbf{x}}_k, FP_kF^T). \tag{17}$$

By property (14) (here $G\mathbf{u}_k$ is a constant vector),

$$F\mathbf{x}_k + G\mathbf{u}_k \sim N(F\hat{\mathbf{x}}_k + G\mathbf{u}_k, FP_kF^T). \tag{18}$$

Further by property (15),

$$\mathbf{x}_{k+1} = (F\mathbf{x}_k + G\mathbf{u}_k) + \mathbf{w}_k \sim N(F\hat{\mathbf{x}}_k + G\mathbf{u}_k, FP_kF^T + Q). \tag{19}$$

Thus if we denote the estimate of $\mathbf{x}_{k+1}$ (after the process but before the observation) as

$$\mathbf{x}_{k+1} \sim N(\bar{\mathbf{x}}_{k+1}, \bar{P}_{k+1}), \tag{20}$$

then **the prediction formula** is

$$\begin{aligned} \bar{\mathbf{x}}_{k+1} &= F\hat{\mathbf{x}}_k + G\mathbf{u}_k, \\ \bar{P}_{k+1} &= FP_kF^T + Q. \end{aligned} \tag{21}$$

This is the formula (5) and (6).

# 6 Kalman Filter update

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

The observation model is

$$\mathbf{z}_{k+1} = H\mathbf{x}_{k+1} + \mathbf{v}_{k+1}. \tag{22}$$

where $H$ is a constant matrix, $\mathbf{z}_{k+1}$ is the observation value at time $k+1$ (constant) and $\mathbf{v}_{k+1}$ is the zero-mean Gaussian observation noise with variance $R$. That is $\mathbf{v}_{k+1} \sim N(0, R)$. It is also assumed that $\mathbf{v}_{k+1}$ is independent with $\mathbf{x}_{k+1}$.

## 6.1 A special case: $H$ is square and invertible

When $H$ is square and invertible, by the observation model (22),

$$\mathbf{x}_{k+1} = -H^{-1}\mathbf{v}_{k+1} + H^{-1}\mathbf{z}_{k+1}. \tag{23}$$

By property (13) (choosing $F = -H^{-1}$),

$$-H^{-1}\mathbf{v}_{k+1} \sim N(0, H^{-1}RH^{-T}). \tag{24}$$

Thus by property (14) (add a constant vector $H^{-1}\mathbf{z}_{k+1}$),

$$\mathbf{x}_{k+1} \sim N(H^{-1}\mathbf{z}_{k+1}, H^{-1}RH^{-T}). \tag{25}$$

The prior information about $\mathbf{x}_{k+1}$ is given by (20) (after the prediction but before the update). So we have two pieces information about $\mathbf{x}_{k+1}$ — information from observation (25) and prior information (20).

According to the definition of information matrix of a Gaussian distribution (see (12) in Section 3), the information matrix (about $\mathbf{x}_{k+1}$) contained in (20) is

$$I_{prior} = \bar{P}_{k+1}^{-1}, \tag{26}$$

while the information matrix (about $\mathbf{x}_{k+1}$) contained in (25) is

$$I_{obs} = (H^{-1}RH^{-T})^{-1} = H^T R^{-1} H. \tag{27}$$

**The total information (about $\mathbf{x}_{k+1}$) after the observation should be the sum of the two**, namely,

$$I_{total} = I_{prior} + I_{obs} = \bar{P}_{k+1}^{-1} + H^T R^{-1} H. \tag{28}$$

**The new mean value is the weighted sum of the mean values of the two Gaussian distributions (25) and (20). The weights are decided by the proportion of information contained in each of the Gaussian distributions (as compared with the total information)**. That is,

$$\begin{aligned} \hat{\mathbf{x}}_{k+1} &= I_{total}^{-1} I_{prior} \bar{\mathbf{x}}_{k+1} + I_{total}^{-1} I_{obs} (H^{-1}\mathbf{z}_{k+1}) \\ &= I_{total}^{-1} I_{prior} \bar{\mathbf{x}}_{k+1} + I_{total}^{-1} H^T R^{-1} \mathbf{z}_{k+1}. \end{aligned} \tag{29}$$

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

$$P_{k+1} = I_{total}^{-1}. \tag{30}$$

So, the final estimate on $\mathbf{x}_{k+1}$ (after the prediction and update) is

$$\mathbf{x}_{k+1} \sim N(\hat{\mathbf{x}}_{k+1}, P_{k+1}) \tag{31}$$

where $\hat{\mathbf{x}}_{k+1}$ and $P_{k+1}$ are given in the above equations (29) and (30).

The update equations (29) and (30) can also be expressed as

$$\begin{aligned} \hat{\mathbf{x}}_{k+1} &= I_{total}^{-1}(I_{prior} + I_{obs} - I_{obs})\bar{\mathbf{x}}_{k+1} + I_{total}^{-1} H^T R^{-1} \mathbf{z}_{k+1} \\ &= I_{total}^{-1}(I_{total} - I_{obs})\bar{\mathbf{x}}_{k+1} + I_{total}^{-1} H^T R^{-1} \mathbf{z}_{k+1} \\ &= \bar{\mathbf{x}}_{k+1} + I_{total}^{-1} H^T R^{-1} (\mathbf{z}_{k+1} - H\bar{\mathbf{x}}_{k+1}) \end{aligned} \tag{32}$$

and

$$P_{k+1} = I_{total}^{-1} = (\bar{P}_{k+1}^{-1} + H^T R^{-1} H)^{-1}. \tag{33}$$

By the matrix inversion lemma in Section 7, in particular equation (44),

$$\begin{aligned} P_{k+1} &= I_{total}^{-1} \\ &= (\bar{P}_{k+1}^{-1} + H^T R^{-1} H)^{-1} \\ &= \bar{P}_{k+1} - \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1} H \bar{P}_{k+1}. \end{aligned} \tag{34}$$

Denote $S = R + H\bar{P}_{k+1}H^T$, then

$$
\begin{aligned}
P_{k+1} &= \bar{P}_{k+1} - \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1} H \bar{P}_{k+1} \\
&= \bar{P}_{k+1} - (\bar{P}_{k+1} H^T S^{-1}) S (S^{-1} H \bar{P}_{k+1}) \\
&= \bar{P}_{k+1} - K S K^T.
\end{aligned} \tag{35}
$$

where $K = \bar{P}_{k+1} H^T S^{-1}$.

This is the formula (8).

Now from (34),

$$
\begin{aligned}
I_{total}^{-1} H^T R^{-1} &= (\bar{P}_{k+1} - \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1} H \bar{P}_{k+1}) H^T R^{-1} \\
&= \bar{P}_{k+1} H^T R^{-1} - \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1} H \bar{P}_{k+1} H^T R^{-1} \\
&= \bar{P}_{k+1} H^T R^{-1} - \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1} [(R + H\bar{P}_{k+1} H^T) - R] R^{-1} \\
&= \bar{P}_{k+1} H^T R^{-1} - \bar{P}_{k+1} H^T R^{-1} + \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1} R R^{-1} \\
&= \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1}.
\end{aligned} \tag{36}
$$

Thus from (32),

$$
\begin{aligned}
\hat{\mathbf{x}}_{k+1} &= \bar{\mathbf{x}}_{k+1} + I_{total}^{-1} H^T R^{-1} (\mathbf{z}_{k+1} - H\bar{\mathbf{x}}_{k+1}) \\
&= \bar{\mathbf{x}}_{k+1} + \bar{P}_{k+1} H^T (R + H\bar{P}_{k+1} H^T)^{-1} (\mathbf{z}_{k+1} - H\bar{\mathbf{x}}_{k+1}) \\
&= \bar{\mathbf{x}}_{k+1} + \bar{P}_{k+1} H^T S^{-1} (\mathbf{z}_{k+1} - H\bar{\mathbf{x}}_{k+1}) \\
&= \bar{\mathbf{x}}_{k+1} + K (\mathbf{z}_{k+1} - H\bar{\mathbf{x}}_{k+1})
\end{aligned} \tag{37}
$$

This is the formula (7).

## 6.2 The general case: $H$ is arbitrary

When $H$ is arbitrary, $H^{-1}$ may not exist but we still have the formula of the information matrix from observation

$$
I_{obs} = H^T R^{-1} H. \tag{38}
$$

Here $I_{obs}$ might not be full rank (the singularity of $I_{obs}$ means that the observation only contain information about part of the state $\mathbf{x}_{k+1}$, there is no information about some part of the vector $\mathbf{x}_{k+1}$ in the observation). However, the total information $I_{total}$ is invertible (full rank).

Also, we still have the new mean value as the weighted sum

$$
\hat{\mathbf{x}}_{k+1} = I_{total}^{-1} I_{prior} \bar{\mathbf{x}}_{k+1} + I_{total}^{-1} H^T R^{-1} \mathbf{z}_{k+1}. \tag{39}
$$

Now all the formula after (27) and (29) are true since $H^{-1}$ is not required.

**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 extended Kalman Filter (EKF), then please read Part III.

# 7 Matrix Inversion Lemma

The following matrix inversion lemma is very useful and can be found in many textbooks about matrices or Kalman Filter.

**Lemma 7.1** *Suppose that the partitioned matrix*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

*is invertible and that the inverse is conformably partitioned as*

$$M^{-1} = \begin{bmatrix} X & Y \\ U & V \end{bmatrix}, \tag{40}$$

*where $A, D, X$ and $V$ are square matrices. If $A$ is invertible, then*

$$\begin{aligned} X &= A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}, \\ Y &= -A^{-1}B(D - CA^{-1}B)^{-1}, \\ U &= -(D - CA^{-1}B)^{-1}CA^{-1}, \\ V &= (D - CA^{-1}B)^{-1}. \end{aligned} \tag{41}$$

*If $D$ is invertible, then*

$$\begin{aligned} X &= (A - BD^{-1}C)^{-1}, \\ Y &= -(A - BD^{-1}C)^{-1}BD^{-1}, \\ U &= -D^{-1}C(A - BD^{-1}C)^{-1}, \\ V &= D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}. \end{aligned} \tag{42}$$

*Thus if both $A$ and $D$ are invertible,*

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}. \tag{43}$$

When $B = C^T$, equation (43) can be written as (substituting $D$ by $-D$)

$$(A + C^T D^{-1}C)^{-1} = A^{-1} - A^{-1}C^T(D + CA^{-1}C^T)^{-1}CA^{-1}. \tag{44}$$