

# Summary of Kalman Filter and Extend Kalman Filter

Yanjun Cao

January 30, 2018

## Abstract

This article will explain the filters from different points of view, probabilistic point of view, control point of view and HMM point of view in machine learning.

## 1 Introduction

### 1.1 Why use Kalman filter?

- State can not be measured directly.
- Measurement has noise and bias, and they are random in one moment.

So if distribution of noise is known, we can use probability theory to filter out the noise.

### 1.2 How does it work?

Kalman filter is a iterative process. With the information of noise distribution, the system can estimate the error based on the estimation in last step.

Separated into predicate and update process. Predicate is the state predication based on previous measurements. Then update process happens when the measurement now comes out. Update combine the predication with this step measurement and then give the final estimate result.

### 1.3 What are needed?

- System process function with process noise distribution.
- Measurement function with measurement noise distribution.

## 2 Derivative

1. Machine learning derive (Yida Xu) In machine learning, three types of problem for **Hidden Markov Model** correspond to three kinds of information want to acquired, observed information, state (latent variable), or parameters. For **observed sequence**, which is measurement, is to find out  $P(y(0), y(1), \dots, y(L-1))$  for a designed measurement sequence  $\{y(0), y(1), \dots, y(L-1)\}$ .

For state, latent variables, four sub-problems: **smoothing, filtering prediction and decoding**. Smoothing is to find state  $(s_m, m < t)$  for the measurement in middle of a given measurement sequence and filtering is for the latest state  $(s_t)$  of the measurement sequence. Prediction is to find next state  $(s_{t+1})$ . And decoding problem is to find a series of states  $P(s_1, s_2, \dots, s_t)$ . The last one is to estimate the parameters of the model system.

Concept:

State transmit with Gaussian noise matrix (Model)

latent variable(things that can not be measured directly)

Two marginal distributions, calculate joint distribution

State transit function:  $x_t = Ax_{t-1} + w_t, w_t \sim N(0, Q_t)$

Observation/Measurement function:  $y_t = Cx_t + v_t, v_t \sim N(0, R_t)$

These two function decide system model. Now we ignore the control function, which can be added easily.

Purpose:  $P(x_t|y_{1:t})$

Add  $x_{t-1}$  into the estimation to make use of model:

$$\begin{aligned}
 P(x_t|y_{1:t}) &= \int_{x_{t-1}} P(x_t, x_{t-1}|y_{1:t}) dx_{t-1} && \text{add } x_{t-1} \\
 &= \int_{x_{t-1}} P(x_t|x_{t-1}, y_{1:t}) * P(x_{t-1}|y_{1:t}) dx_{t-1} && \text{joint-} > \text{conditional} \\
 &= \int_{x_{t-1}} P(x_t|x_{t-1}) * P(x_{t-1}|y_{1:t}) dx_{t-1} && \text{HMM property} \\
 &= \int_{x_{t-1}} P(x_t|x_{t-1}) * P(x_{t-1}|y_t, y_{1:t-1}) dx_{t-1} && \text{seperate } y_t \\
 &= \int_{x_{t-1}} P(x_t|x_{t-1}) * \frac{P(y_t|x_{t-1}, y_{1:t-1}) * P(x_{t-1}|y_{1:t-1})}{P(y_t|y_{1:t-1})} dx_{t-1} && \text{bayes rule} \\
 &= \eta * \int_{x_{t-1}} P(x_t|x_{t-1}) * P(y_t|x_{t-1}) * P(x_{t-1}|y_{1:t-1}) dx_{t-1} && \text{HMM property}
 \end{aligned}$$

This is for understanding, not derivative.

$$\begin{aligned}
P(x_t|y_{1:t}) &= P(x_t|y_t, y_{1:t-1}) && \text{seperate } y_t \\
&= \eta P(y_t|x_t, y_{1:t-1}) * P(x_t|y_{1:t-1}) && \text{bayes rule} \\
&= \eta P(y_t|x_t) * P(x_t|y_{1:t-1}) && \text{hmm property} \\
&= \eta P(y_t|x_t) \int_{x_{t-1}} P(x_t, x_{t-1}|y_{1:t-1}) dx_{t-1} && \text{add } x_{t-1} \\
&= \eta P(y_t|x_t) \int_{x_{t-1}} P(x_t|x_{t-1}, y_{1:t-1}) * P(x_{t-1}|y_{1:t-1}) dx_{t-1} && \text{joint to conditional} \\
&= \eta P(y_t|x_t) \int_{x_{t-1}} P(x_t|x_{t-1}) * P(x_{t-1}|y_{1:t-1}) dx_{t-1} && \text{hmm property} \\
&= \text{Now it is recursive, all are gaussian}
\end{aligned}$$

Now to derivative.

Suppose we have information about step  $t-1$ .  $P(x_{t-1}|y_{1:t-1}) \sim N(\bar{u}_{t-1}, \bar{\sigma}_{t-1})$

Firstly to derive  $P(x_t|y_{1:t-1})$ , and  $P(y_t|y_{1:t-1})$  from system model formulas based on pre-step informations. Secondly to write out  $P(x_t, y_t|y_{1:t-1})$ . At last write  $P(x_t|y_t, y_{1:t-1})$  from properties of the multivariate normal.

From transit formula,  $P(x_t|x_{t-1}) \sim N(Ax_{t-1}, w_t)$

From observation formula,  $P(y_t|y_{1:t-1}) \sim N(HA\bar{u}_{t-1}, HA\bar{\sigma}_{t-1} + Hw + v)$

Then derive the joint probability  $P(x_t, y_t)$ , then can write the  $P(x_t|y_t)$

2. Robotics (Cyrill)
3. Control theory (Matlab)