Optimization and Optimal Control

Vahid Hassani

OsloMet

October 15, 2019

Kalman Filter

Formulating the Kalman filter in stochastic continuous-time framework needs Itō calculus and is out of scope of this course. Hence, we only touch the basics of Kalman filtering in stochastic discrete-time framework.

Motivation

- Given a Discrete-time, linear, time varying plant with random initial state and driven by white plant noise.
- Given noisy measurements of linear combinations of the plant state variables
- Determine the "best" estimate of the plant state variables.

Kalman Filter

Formulating the Kalman filter in stochastic continuous-time framework needs Itō calculus and is out of scope of this course. Hence, we only touch the basics of Kalman filtering in stochastic discrete-time framework.

Motivation

- Given a Discrete-time, linear, time varying plant with random initial state and driven by white plant noise.
- Given noisy measurements of linear combinations of the plant state variables
- Determine the “best” estimate of the plant state variables.

Kalman Filter (Contd.)

Problem Formulation

- $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$ State dynamics
- $y(t+1) = C(t+1)x(t+1) + \theta(t+1)$ Measurement equation

  $t = 0, 1, 2, \ldots$

- $x(t) \in \mathbb{R}^n$, State vector (stochastic sequence non-white)
- $u(t) \in \mathbb{R}^m$, Deterministic input sequence
- $\xi(t) \in \mathbb{R}^p$, White plant noise sequence
- $\theta(t) \in \mathbb{R}^r$, White measurement noise sequence
- $y(t) \in \mathbb{R}^r$, Measurement vector

Kalman Filter (Contd.)

Probabilistic Information

- $E\{x(0)\} = x_0, \quad cov[x(0); x(0)] = \Sigma_0 = \Sigma_0^T \succeq 0$
  Initial state $x(0)$ is gaussian
- $E\{\xi(t)\} = 0, \quad cov[\xi(t); \xi(\tau)] = \Xi(t)\delta_{t\tau} = \Xi(t)^T\delta_{t\tau} \succeq 0$
  Plant noise $\xi(t)$ is gaussian discret white noise sequence
- $E\{\theta(t)\} = 0, \quad cov[\theta(t); \theta(\tau)] = \Theta(t)\delta_{t\tau} = \Theta(t)^T\delta_{t\tau} \succeq 0$
  Measurment noise $\theta(t)$ is gaussian discret white noise sequence
- $x(0)$, $\xi(t)$, and $\theta(\tau)$ are independent for all t, τ

Kalman Filter (Contd.)

Definition of the Filtering Problem

- Let t denote present value of time.
- Given the sequence of the past inputs $U(t-1) := \{u(0), u(1), \ldots, u(t-1)\}$
- Given the sequence of the past measurement $Y(t) := \{y(1), y(2), \ldots, y(t)\}$
- Determine a “good” estimate of $x(t)$.

Kalman Filter (Contd.)

Good to remember that:

The linearity of

a) the state equation

b) the measurement equation

and the gaussian nature of

a) the initial state, $x(0)$

b) the plant white noise $\xi(t)$

c) the measurement white noise $\theta(t)$

imply that $p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is gaussian!

$p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is called conditional density function.

Kalman Filter (Contd.)

Good to remember that:

The linearity of

a) the state equation

b) the measurement equation

and the gaussian nature of

a) the initial state, $x(0)$

b) the plant white noise $\xi(t)$

c) the measurement white noise $\theta(t)$

imply that $p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is gaussian!

$p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is called conditional density function.

Kalman Filter (Contd.)

Since $p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is Gaussian:

The conditional state density function is uniquely characterized by

(i) the conditional mean

(ii) the conditional covariance

where

$$\hat{x}(t \mid t) := E\{x(t) \mid Y(t), U(t-1)\} = \int x(t)p\big(x(t) \mid Y(t), U(t-1)\big)dx(t)$$

Since the conditional density function is gaussian, all the reasonable estimates(mean, median, most probable) are the same.

$$\Sigma(t \mid t) := cov[x(t); x(t) \mid Y(t), U(t-1)])$$

$$= \int [x(t) - \hat{x}(t \mid t)][x(t) - \hat{x}(t \mid t)]^T p\big(x(t) \mid Y(t), U(t-1)\big)dx(t)$$

Kalman Filter (Contd.)

Time Structure of The Problem:

The development has an inductive flavor. The basic process is as follows:

1. Assume that all relevant quantities are available at time t, i.e. $Y(t),\ U(t-1)$. Then:
   - "Nature" applies $\xi(t)$
   - We apply $u(t)$
   - The system moves from state $x(t)$ to state $x(t+1)$ according to $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$
   - We make a measurement $y(t+1)$ based on $y(t+1) = C(t+1)x(t+1) + \theta(t+1)$
2. We want now to make an estimate of $x(t+1)$ based on the expanded information $Y(t+1) = \{Y(t), y(t+1)\}$ and $U(t) = \{U(t-1), u(t)\}$ (the "newest" measurement and control is added to Y and U, respectively).

Kalman Filter (Contd.)

The estimation Process is divided into two parts:

1. PREDICT CYCLE

   What can we say about $x(t+1)$ ***before*** we make the measurement $y(t+1)$.

   For predict cycle we have:

   - $U(t) = \{u(0), u(1), \ldots, u(t-1), u(t)\}$
   - $Y(t) = \{y(1), y(2), \ldots, y(t-1), y(t)\}$

2. UPDATE CYCLE

   How can we improve our information about $x(t+1)$ ***after*** we make the measurement $y(t+1)$.

   For update cycle we have:

   - $U(t) = \{u(0), u(1), \ldots, u(t-1), u(t)\}$
   - $Y(t+1) = \{y(1), y(2), \ldots, y(t-1), y(t), y(t+1)\}$

Kalman Filter (Contd.)

The estimation Process is divided into two parts:

1. PREDICT CYCLE

   What can we say about $x(t+1)$ ***before*** we make the measurement $y(t+1)$.

   For predict cycle we have:

   - $U(t) = \{u(0), u(1), \ldots, u(t-1), u(t)\}$
   - $Y(t) = \{y(1), y(2), \ldots, y(t-1), y(t)\}$

2. UPDATE CYCLE

   How can we improve our information about $x(t+1)$ ***after*** we make the measurement $y(t+1)$.

   For update cycle we have:

   - $U(t) = \{u(0), u(1), \ldots, u(t-1), u(t)\}$
   - $Y(t+1) = \{y(1), y(2), \ldots, y(t-1), y(t), y(t+1)\}$

Kalman Filter (Contd.)

What do we need to process the measurement $y(t+1)$?

The key quantity that needs to be evaluated is $p(x(t+1) \mid Y(t+1), U(t))$.

Probability density function of $x(t+1)$ given $Y(t)$, $U(t)$, i.e. $p(x(t+1) \mid Y(t), U(t))$, can be seen as the "prior" information (before making the measurement $y(t+1)$).

Then the Bayes rule ($p(A \mid B) = \frac{p(B|A)p(A)}{p(B)}$) requires:

$$p(x(t+1) \mid Y(t+1), U(t)) = \frac{p(y(t+1) \mid x(t+1), Y(t), U(t))\, p(x(t+1) \mid Y(t), U(t))}{p(y(t+1) \mid Y(t), U(t))}$$

Kalman Filter (Contd.)

What do we need to process the measurement $y(t+1)$?

The key quantity that needs to be evaluated is $p\big(x(t+1) \mid Y(t+1), U(t)\big)$.

Probability density function of $x(t+1)$ given $Y(t)$, $U(t)$, i.e. $p\big(x(t+1) \mid Y(t), U(t)\big)$, can be seen as the "prior" information (before making the measurement $y(t+1)$).

Then the Bayes rule ($p(A \mid B) = \frac{p(B|A)p(A)}{p(B)}$) requires:

$$p\big(x(t+1) \mid Y(t+1), U(t)\big) = \frac{p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)\, p\big(x(t+1) \mid Y(t), U(t)\big)}{p\big(y(t+1) \mid Y(t), U(t)\big)}$$

Kalman Filter (Contd.)

What do we need to process the measurement $y(t+1)$?

The key quantity that needs to be evaluated is
$p\big(x(t+1) \mid Y(t+1), U(t)\big)$.

Probability density function of $x(t+1)$ given $Y(t)$, $U(t)$, i.e. $p\big(x(t+1) \mid Y(t), U(t)\big)$, can be seen as the "prior" information (before making the measurement $y(t+1)$).

Then the Bayes rule ($p(A \mid B) = \frac{p(B|A)p(A)}{p(B)}$) requires:

$$p\big(x(t+1) \mid Y(t+1), U(t)\big) = \frac{p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)\, p\big(x(t+1) \mid Y(t), U(t)\big)}{p\big(y(t+1) \mid Y(t), U(t)\big)}$$

Kalman Filter (Contd.)

What do we need to process the measurement $y(t+1)$?

The key quantity that needs to be evaluated is $p\big(x(t+1) \mid Y(t+1), U(t)\big)$.

Probability density function of $x(t+1)$ given $Y(t)$, $U(t)$, i.e. $p\big(x(t+1) \mid Y(t), U(t)\big)$, can be seen as the “prior” information (before making the measurement $y(t+1)$).

Then the Bayes rule ($p(A \mid B) = \frac{p(B|A)p(A)}{p(B)}$) requires:

$$p\big(x(t+1) \mid Y(t+1), U(t)\big) = \frac{p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)\, p\big(x(t+1) \mid Y(t), U(t)\big)}{p\big(y(t+1) \mid Y(t), U(t)\big)}$$

Kalman Filter (Contd.)

$$p\big(x(t+1) \mid Y(t+1), U(t)\big) = \frac{p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)p\big(x(t+1) \mid Y(t), U(t)\big)}{p\big(y(t+1) \mid Y(t), U(t)\big)}$$

Let us take a look on these three terms:

- $p\big(x(t+1) \mid Y(t), U(t)\big)$
- $p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)$
- $p\big(y(t+1) \mid Y(t), U(t)\big)$

Kalman Filter (Contd.)

How to compute $p(x(t+1) \mid Y(t), U(t))$

Recall that $p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is Gaussian and

$$\hat{x}(t \mid t) := E\big\{x(t) \mid Y(t), U(t-1)\big\} = \int x(t)p\big(x(t) \mid Y(t), U(t-1)\big)dx(t)$$

and $\Sigma(t \mid t) := cov[x(t); x(t) \mid Y(t), U(t-1)])$ are known.

Remember system dynamics $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$.

$[x(t) \mid Y(t), U(t-1)]$ and $\xi(t)$ are gaussaian and independent; $u(t)$ is deterministic and known.

so $[x(t+1) \mid Y(t), U(t)]$ is gaussaian.

Kalman Filter (Contd.)

How to compute $p(x(t+1) \mid Y(t), U(t))$

Recall that $p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is Gaussian and

$$\hat{x}(t \mid t) := E\{x(t) \mid Y(t), U(t-1)\} = \int x(t)p\big(x(t) \mid Y(t), U(t-1)\big)dx(t)$$

and $\Sigma(t \mid t) := cov[x(t); x(t) \mid Y(t), U(t-1)])$ are known.

Remember system dynamics $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$.

$[x(t) \mid Y(t), U(t-1)]$ and $\xi(t)$ are gaussaian and independent; $u(t)$ is deterministic and known.

so $[x(t+1) \mid Y(t), U(t)]$ is gaussaian.

Kalman Filter (Contd.)

How to compute $p(x(t+1) \mid Y(t), U(t))$

Recall that $p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is Gaussian and

$$\hat{x}(t \mid t) := E\{x(t) \mid Y(t), U(t-1)\} = \int x(t)p(x(t) \mid Y(t), U(t-1))dx(t)$$

and $\Sigma(t \mid t) := cov[x(t); x(t) \mid Y(t), U(t-1)])$ are known.

Remember system dynamics $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$.

$[x(t) \mid Y(t), U(t-1)]$ and $\xi(t)$ are gaussaian and independent; $u(t)$ is deterministic and known.

so $[x(t+1) \mid Y(t), U(t)]$ is gaussaian.

Kalman Filter (Contd.)

How to compute $p(x(t+1) \mid Y(t), U(t))$

Recall that $p\Big(x(t) \mid Y(t), U(t-1)\Big)$ is Gaussian and

$$\hat{x}(t \mid t) := E\{x(t) \mid Y(t), U(t-1)\} = \int x(t) p(x(t) \mid Y(t), U(t-1))\, dx(t)$$

and $\Sigma(t \mid t) := cov[x(t); x(t) \mid Y(t), U(t-1)])$ are known.

Remember system dynamics $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$.

$[x(t) \mid Y(t), U(t-1)]$ and $\xi(t)$ are gaussaian and independent; $u(t)$ is deterministic and known.

so $[x(t+1) \mid Y(t), U(t)]$ is gaussaian.

Kalman Filter (Contd.)

PREDICT CYCLE or How to compute $p(x(t+1) \mid Y(t), U(t))$

$$\hat{x}(t+1 \mid t) := E\{x(t) \mid Y(t), U(t)\}$$
$$\Sigma(t+1 \mid t) := cov[x(t+1); x(t+1) \mid Y(t), U(t)])$$

How to calculate? (recall $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$)

$$\boxed{\hat{x}(t+1 \mid t) = A(t)\,\hat{x}(t \mid t) + B(t)u(t)}$$

$$\boxed{\Sigma(t+1 \mid t) = A(t)\,\Sigma(t \mid t)\,A(t)^T + L(t)\,\Xi(t)\,L(t)^T}$$

Kalman Filter (Contd.)

PREDICT CYCLE or How to compute $p\big(x(t+1) \mid Y(t), U(t)\big)$

$$\hat{x}(t+1 \mid t) := E\big\{x(t) \mid Y(t), U(t)\big\}$$
$$\Sigma(t+1 \mid t) := cov[x(t+1); x(t+1) \mid Y(t), U(t)]$$

How to calculate? (recall $x(t+1) = A(t)x(t) + B(t)u(t) + L(t)\xi(t)$)

$$\boxed{\hat{x}(t+1 \mid t) = A(t)\,\hat{x}(t \mid t) + B(t)u(t)}$$

$$\boxed{\Sigma(t+1 \mid t) = A(t)\,\Sigma(t \mid t)\,A(t)^T + L(t)\,\Xi(t)\,L(t)^T}$$

Kalman Filter (Contd.)

How to compute $p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)$

Recall that Measurement equation is in the form of $y(t+1) = C(t+1)x(t+1) + \theta(t+1)$. In calculation of

$p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)$ the term $x(t+1)$ is given; so

$$p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big) \;=\; p\big(y(t+1) \mid x(t+1)\big)$$

Since $\theta(t+1)$ is gaussaian, so is $p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)$ and

$$E\big\{y(t+1) \mid x(t+1), Y(t), U(t)\big\} = C(t+1)x(t+1)$$
$$cov[y(t+1); y(t+1) \mid x(t+1), Y(t), U(t)]) = \Theta(t+1)$$

Kalman Filter (Contd.)

How to compute $p\big(y(t+1) \mid Y(t), U(t)\big)$

Recall the measurement equation $y(t+1) = C(t+1)x(t+1) + \theta(t+1)$ and the fact that $x(t+1)$ and $\theta(t+1)$ are independent.

So $p\big(y(t+1) \mid Y(t), U(t)\big)$ is gaussian.

Recall that

$$\hat{x}(t+1 \mid t) := E\big\{x(t+1) \mid Y(t), U(t)\big\} = A(t)\,\hat{x}(t \mid t) + B(t)u(t)$$

$$\Sigma(t+1 \mid t) := cov[x(t+1); x(t+1) \mid Y(t), U(t)]) = A(t)\,\Sigma(t \mid t)\,A(t)^T + L(t)\,\Xi(t)\,L(t)^T$$

It follows that

$$E\big\{y(t+1) \mid Y(t), U(t)\big\} = C(t+1)\,\hat{x}(t+1 \mid t)$$

$$cov[y(t+1); y(t+1) \mid Y(t), U(t)]) = C(t+1)\,\Sigma(t+1 \mid t)\,C(t+1)^T + \Theta(t+1)$$

Kalman Filter (Contd.)

Remember that we wanted to compute

$$p\big(x(t+1) \mid Y(t+1), U(t)\big) = \frac{p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)p\big(x(t+1) \mid Y(t), U(t)\big)}{p\big(y(t+1) \mid Y(t), U(t)\big)}$$

And we have already calculated all the followings

- $p\big(x(t+1) \mid Y(t), U(t)\big)$
- $p\big(y(t+1) \mid x(t+1), Y(t), U(t)\big)$
- $p\big(y(t+1) \mid Y(t), U(t)\big)$

Kalman Filter (Contd.)

Update CYCLE or How to compute $p(x(t+1) \mid Y(t+1), U(t))$

$$\hat{x}(t+1 \mid t+1) := E\{x(t+1) \mid Y(t+1), U(t)\}$$
$$\Sigma(t+1 \mid t+1) := cov[x(t+1); x(t+1) \mid Y(t+1), U(t)])$$

How to calculate? (Use the Bayes rule and our findings)

$$\boxed{\hat{x}(t+1 \mid t+1) = \hat{x}(t+1 \mid t) + \Sigma(t+1 \mid t+1)C^T(t+1)\Theta^{-1}(t+1)[y(t+1) - c(t+1)\hat{x}(t+1 \mid t)]}$$

$$\boxed{\Sigma(t+1 \mid t+1) = \Sigma(t+1 \mid t) - \Sigma(t+1 \mid t)C^T(t+1)[C(t+1)\Sigma(t+1 \mid t)C^T(t+1) + \Theta(t+1)]^{-1}C(t+1)\Sigma(t+1 \mid t)}$$

Kalman Filter (Contd.)

Update CYCLE or How to compute $p\big(x(t+1) \mid Y(t+1), U(t)\big)$

$$\hat{x}(t+1 \mid t+1) := E\big\{x(t+1) \mid Y(t+1), U(t)\big\}$$
$$\Sigma(t+1 \mid t+1) := cov[x(t+1); x(t+1) \mid Y(t+1), U(t)])$$

How to calculate? (Use the Bayes rule and our findings)

$$\boxed{\hat{x}(t+1 \mid t+1) = \hat{x}(t+1 \mid t) + \Sigma(t+1 \mid t+1)C^T(t+1)\Theta^{-1}(t+1)\big[y(t+1) - c(t+1)\hat{x}(t+1 \mid t)\big]}$$

$$\boxed{\Sigma(t+1 \mid t+1) = \Sigma(t+1 \mid t) - \Sigma(t+1 \mid t)C^T(t+1)\big[C(t+1)\Sigma(t+1 \mid t)C^T(t+1) + \Theta(t+1)\big]^{-1}C(t+1)\Sigma(t+1 \mid t)}$$

Stochastic discrete-Time Kalman Filter (Summary)

OFF-LINE CALCULATION

- Initialization ($t = 0$)
  $\Sigma(0 \mid 0) = cov[x(0); x(0)]$
- Predict Cycle
  $\Sigma(t+1 \mid t) = A(t)\, \Sigma(t \mid t)\, A^T(t) + L(t)\, \Xi(t)\, L^T(t)$
- Update Cycle
  $\Sigma(t+1 \mid t+1) = \Sigma(t+1 \mid t) - \Sigma(t+1 \mid t)C^T(t+1)\left[C(t+1)\Sigma(t+1 \mid t)C^T(t+1) + \Theta(t+1)\right]^{-1} C(t+1)\Sigma(t+1 \mid t)$
- Filter Gain Matrix
  $H(t+1) = \Sigma(t+1 \mid t+1)C^T(t+1)\Theta^{-1}(t+1)$

ON-LINE CALCULATION

- Initialization ($t = 0$)
  $\hat{x}(0 \mid 0) = E\{x(0)\}$
- Predict Cycle
  $\hat{x}(t+1 \mid t) = A(t)\, \hat{x}(t \mid t) + B(t)u(t)$
- Update Cycle
  $\hat{x}(t+1 \mid t+1) = \hat{x}(t+1 \mid t) + H(t+1)\left[y(t+1) - c(t+1)\hat{x}(t+1 \mid t)\right]$

Show that residuals are zero mean white noise!

Stochastic discrete-Time Kalman Filter (Summary)

OFF-LINE CALCULATION

- Initialization ($t = 0$)
  $\Sigma(0 \mid 0) = cov[x(0); x(0)]$
- Predict Cycle
  $\Sigma(t+1 \mid t) = A(t)\,\Sigma(t \mid t)\,A^T(t) + L(t)\,\Xi(t)\,L^T(t)$
- Update Cycle
  $\Sigma(t+1 \mid t+1) = \Sigma(t+1 \mid t) - \Sigma(t+1 \mid t)C^T(t+1)\left[C(t+1)\Sigma(t+1 \mid t)C^T(t+1) + \Theta(t+1)\right]^{-1} C(t+1)\Sigma(t+1 \mid t)$
- Filter Gain Matrix
  $H(t+1) = \Sigma(t+1 \mid t+1)C^T(t+1)\Theta^{-1}(t+1)$

ON-LINE CALCULATION

- Initialization ($t = 0$)
  $\hat{x}(0 \mid 0) = E\{x(0)\}$
- Predict Cycle
  $\hat{x}(t+1 \mid t) = A(t)\,\hat{x}(t \mid t) + B(t)u(t)$
- Update Cycle
  $\hat{x}(t+1 \mid t+1) = \hat{x}(t+1 \mid t) + H(t+1)\left[y(t+1) - c(t+1)\hat{x}(t+1 \mid t)\right]$

Show that residuals are zero mean white noise!

References:

- "Optimal Filtering", Brian D. O. Anderson and John B. Moore, Prentice-Hall, 1979
  Available online at
  `http://users.cecs.anu.edu.au/~john/papers/BOOK/B02.PDF`