This set of lectures provides a brief introduction to Kalman filtering, following the treatment in Friedland.

Reading:

- Friedland, Chapter 11

1 Introduction

![Block diagram of a basic feedback loop.](image)

Figure 1: Block diagram of a basic feedback loop.

2 Linear Quadratic Estimators

Consider a stochastic system
\[
\begin{align*}
\dot{x} &= Ax + Bu + Fv \\
y &= Cx + w
\end{align*}
\]
Assume that the disturbance \(v\) and noise \(w\) are zero-mean and Gaussian (but not necessarily stationary):
\[
\begin{align*}
p(v) &= \frac{1}{\sqrt{2\pi} \sqrt{\det Q}} e^{-\frac{1}{2}v^TQ^{-1}v} \\
p(w) &= \ldots \quad \text{(using R)}
\end{align*}
\]
- multi-variable Gaussian with covariance matrix $Q$
- in scalar case, $Q = \sigma^2$

**Problem statement:** Find the estimate $\hat{x}(t)$ that minimizes the mean square error $E\{(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T\}$ given $\{y(\tau) : 0 \leq \tau \leq t\}$.

**Proposition** $\hat{x}(t) = E\{x(t)|y(\tau), \tau \leq t\}$

- Optimal estimate is just the expectation of the random process $x$ given the constraint of the observed output.
- This is the way Kalman originally formulated the problem.
- Can think of this as a least squares problem: given all previous $y(t)$, find the estimate $\hat{x}$ that satisfies the dynamics and minimizes the square error with the measured data.

**Proof** See text. Basic idea: show that the conditional mean minimizes the mean square error.

**Theorem 1 (Kalman-Bucy, 1961).** The optimal estimator has the form of a linear observer

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x})$$

where $L(t) = P(t)C^T R^{-1}$ and $P(t) = E\{(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T\}$ and satisfies

$$\dot{P} = AP + PA^T - PC^T R^{-1}(t)CP + FQ(t)F^T$$

$P(0) = E\{x(0)x^T(0)\}$

**Proof.** (sketch) The error dynamics are given by

$$\dot{e} = (A - LC)e + \xi \quad \xi = Fv - Lw$$

$$R_\xi = FQF^T + LRL^T$$

The covariance matrix $P_e = P$ for this process satisfies (from last lecture):

$$\dot{P} = (A - LC)P + P(A - LC)^T + FQF^T + LRL^T.$$

We need to find $L$ such that $P(t)$ is as small as possible. Can show that the $L$ that achieves this is given by

$$RL^T = CP \quad \implies \quad L = PC^T R^{-1}$$

(See Friedland, Section 9.4).
1. The Kalman filter has the form of a recursive filter: given $P(t) = E\{e(t)e^T(t)\}$ at time $t$, can compute how the estimate and covariance change. Don’t need to keep track of old values of the output.

2. The Kalman filter gives the estimate $\hat{x}(t)$ and the covariance $P_e(t) \Rightarrow$ you can see how well the error is converging.

3. If the noise is stationary ($Q$, $R$ constant) and if $P$ is stable, then the observer gain is constant:

$$L = PC^TR^{-1} \quad AP + PA^T - PC^TR^{01}CP + FQF^T$$ (algebraic Riccati equation)

This is the problem solved by the lqe command in MATLAB.

4. The Kalman filter extracts the maximum possible information about output data

$$r = y - C\hat{x} = \text{residual or innovations process}$$

Can show that for the Kalman filter, the correlation matrix is

$$R_e(t, s) = W(t)\delta(t - s) \Rightarrow \text{white noise}$$

So the output error has no remaining dynamic information content (see Friedland section 11.5 for complete calculation)

3 Extended Kalman Filters

Consider a nonlinear system

$$\dot{x} = f(x, u, v) \quad x \in \mathbb{R}^n, u \in \mathbb{R}^m$$

$$y = Cx + w \quad v, w \text{ Gaussian white noise processes with}$$

$$\text{covariance matrices } Q \text{ and } R.$$

Nonlinear observer:

$$\dot{\hat{x}} = f(\hat{x}, u, 0) + L(y - C\hat{x})$$

Error dynamics: $e = x - \hat{x}$

$$\dot{e} = f(x, u, v) - f(\hat{x}, u, 0) - LC(x - \hat{x})$$

$$= F(e, \hat{x}, u, v) - LCe \quad F(e, \hat{x}, u, v) = f(e + \hat{x}, u, v) - f(\hat{x}, u, 0)$$

Now linearize around current estimate $\hat{x}$

$$\dot{e} = \frac{\partial F}{\partial e} e + F(0, \hat{x}, u, 0) + \frac{\partial F}{\partial v} v - LCe + \text{h.o.t}$$

$$= \tilde{A}e + \tilde{F}v - LCe$$
where
\[
\tilde{A} = \left. \frac{\partial F}{\partial e} \right|_{(0, \hat{x}, u, 0)} = \left. \frac{\partial f}{\partial x} \right|_{(\hat{x}, u, 0)}
\]
\[
\tilde{F} = \left. \frac{\partial F}{\partial v} \right|_{(0, \hat{x}, u, 0)} = \left. \frac{\partial f}{\partial v} \right|_{(\hat{x}, u, 0)}
\]
\[
\begin{aligned}
&\text{Depend on current} \\
&\text{estimate } \hat{x}
\end{aligned}
\]

Idea: design observer for the linearized system around current estimate
\[
\dot{\hat{x}} = f(\hat{x}, u, 0) + L(y - C\hat{x})
\]
\[
\dot{P} = (\tilde{A} - LC)P + P(\tilde{A} - LC)^T + \tilde{F}Q\tilde{F}^T + LRL^T
\]
\[
P(t_0) = E\{x(t_0)x^T(t_0)\}
\]

This is called the (Schmidt) extended Kalman filter (EKF)

Remarks:

1. Can’t prove very much about EKF due to nonlinear terms
2. In applications, works very well. One of the most used forms of the Kalman filter

**Application: parameter ID**

Consider a linear system with unknown parameters $\xi$
\[
\begin{align*}
\dot{x} &= A(\xi)x + B(\xi)u + Fv \\
y &= C(\xi)x + w
\end{align*}
\]
\[
\xi \in \mathbb{R}^p
\]

Parameter ID problem: given $u(t)$ and $y(t)$, estimate the value of the parameters $\xi$.

One approach: treat $\xi$ as unknown state
\[
\begin{align*}
\dot{x} &= A(\xi)x + B(\xi)u + Fv \\
\dot{\xi} &= 0
\end{align*}
\]
\[
\begin{aligned}
&\text{→} \\
&\frac{d}{dt} \begin{bmatrix} x \\ \xi \end{bmatrix} = \begin{bmatrix} A(\xi) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \xi \end{bmatrix} + \begin{bmatrix} B(\xi) \\ 0 \end{bmatrix} u + \begin{bmatrix} F \\ 0 \end{bmatrix} v \\
y &= C(\xi)x + w
\end{aligned}
\]

Now use EKF to estimate $x$ and $\xi$ \implies determine unknown parameters $\xi \in \mathbb{R}^p$.

Remark: need various observability conditions on augmented system in order for this to work.
4 LQG Control

Return to the original “$H_2$” control problem

\[
\dot{x} = Ax + Bu + Fv \\
y = Cx + w \\
v, w \text{ Gaussian white noise with covariance } R_v, R_w
\]

Stochastic control problem: find $C(s)$ to minimize

\[
J = E \left\{ \int_0^\infty \left[ (y - r)^T Q (y - r)^T + u^T R u \right] dt \right\}
\]

Assume for simplicity that $r = 0$ (otherwise, translate the state accordingly).

**Theorem 2.** The optimal controller has the form

\[
\dot{\hat{x}} = A \hat{x} + Bu + L(y - C \hat{x}) \\
u = K(\hat{x} - x_d)
\]

where $L$ is the optimal observer gain ignoring the controller and $K$ is the optimal controller gain ignoring the noise.

This is called the separation principle (for $H_2$ control).

5 Sensor Fusion