Chapter 7

Kalman Filter

In this chapter, we study Kalman Filtering, which is a celebrated "optimal" observer/filter/prediction design method. In many ways, Kalman filter is the dual of LQ control. We shall develop Kalman filter in this way.

Recall that for time invariant systems, the design method for state feedback can be used to design a state estimator (observer).

Here, the **State Estimator** design problem is to choose L in:

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y(t) - C\hat{x}) \Rightarrow \qquad \dot{\tilde{x}} = (A - LC)\tilde{x}$$

so that the observer error dynamics is stable. This is ensured if $A - LC$ is stable.

The **related State feedback Problem** is to choose K in

$$\begin{aligned} \dot{x} &= A^T x + C^T u; \quad u = -Kx \\ \Rightarrow \dot{x} &= (A^T - C^T K)x. \end{aligned}$$

$A^T - C^T K$ is stable. By choosing $L = K^T$ for the observer, the observer is ensured to be stable.

Sine the K obtained by LQ optimal (infinite horizon) control design is stabilizing as long as some stabilizability and detectability conditions are satisfied, $L = K^T$ can be used as a stabilizing observer gain as well.

The questions we need to ask are:

- How is it done - in the steady state case, and in the time varying (or finite horizon) case?
- In what way is the resulting observer "optimal"?

7.1 Steady State Kalman Filter

Consider first the **Infinite Horizon LQ Control**:

$$\dot{x} = Ax + Bu$$

where R is positive definite, $Q = Q^{\frac{T}{2}} Q^{\frac{1}{2}}$ with (A, B) stabilizable, $(A, Q^{\frac{1}{2}})$ detectable.

We can solve for the positive (semi-) definite P_∞ in the ARE:

$$A^T P_\infty + P_\infty A - P_\infty B R^{-1} B^T P_\infty + Q = 0. \tag{7.1}$$

e.g. via the Hamiltonian matrix approach (using its eigenvectors) or by simply integrating the Riccati equation backwards in time.

We can get the optimal feedback gain: $u = -Kx$ where

$$K = R^{-1}B^T P_\infty$$

Because (A, B) and $(A, Q^{\frac{1}{2}})$ are stabilizable and detectable, $A - BK$ has eigenvalues on the open left half plane.

The stable observer problem is to find L in:

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

so that $A - LC$ is stable.

Let us solve LQ control problem for the dual problem:

$$\dot{x} = A^T x + C^T u \tag{7.2}$$

Transform the Algebraic Riccati Equation: $C^T \to B$, $A^T \to A$.

$$AP_\infty + P_\infty A^T - P_\infty C^T R^{-1} C P_\infty + Q = 0.$$

Stabilizing feedback gain K for (7.2) is:

$$K = L^T = R^{-1}CP_\infty \quad \Rightarrow \quad L = P_\infty C^T R^{-1}$$

This generates the observer:

$$\dot{\hat{x}} = A\hat{x} + Bu + P_\infty C^T R^{-1}(y - C\hat{x})$$

with observer error dynamics:

$$\dot{\tilde{x}} = (A - P_\infty C^T R^{-1} C)\tilde{x}$$

which will be stable, as long as

1. (A^T, C^T) is stabilizable. This is the same as (A, C) being detectable.
2. $(A^T, Q^{\frac{1}{2}})$ is detectable. This is the same as $(A, Q^{\frac{1}{2}})$ being stabilizable.

This design can be easily achieved by merely solving the Algebraic Riccati Equation (ARE). Matlab's `ARE.m` command can be used. Questions we need to ask are:

- In what way is this observer optimal?
- How does finite time LQ extend to observer design?

7.2 Deterministic Kalman Filter

Problem statement:
Suppose we are given the output $y(\tau)$ of a process over the interval $[t_0, t_f]$. It is anticipated that $y(\tau)$ is generated from a process of the form:

$$\begin{aligned} \dot{x} &= A(t)x + B(t)u \\ y &= C(t)x \end{aligned}$$

To account for model uncertainty and measurement noise, the process model used is modified to:

$$\dot{\hat{x}} = A(t)\hat{x} + B(t)u + B_w(t)w$$
$$y = C(t)\hat{x} + v \tag{7.3}$$

The state $\hat{x}(t|t_f)$ is the state estimate the actual process at time t given all the measurements up to time t_f. $\dot{\hat{x}}(t|t_f)$ denotes derivative w.r.t. t. The final time t_f is assumed to be fixed. In (7.3), w and v are the process and measurement noises that are assumed to be unknown. The initial state $\hat{x}(t_0|t_f)$ is also assumed to be unknown.

Choose the initial state $\hat{x}(t_0|t_f)$ and the measurement and process noises $w(\tau|t_f)$, $v(\tau|t_f)$ for $\tau \in [t_0, t_f]$, so that:

1. *The measured output is explained: i.e.*

$$y(\tau) = C(\tau)\hat{x}(\tau|t_f) + v(\tau|t_f)$$

   for all $\tau \in [t_0, t_f]$.

2. *Minimize the following objective function:*

$$J(y(\cdot), \hat{x}(t_0|t_f), w(\cdot|t_f)) = \frac{1}{2}\hat{x}^T(t_0|t_f)S^{-1}\hat{x}(t_0|t_f) +$$
$$+ \frac{1}{2}\int_{t_0}^{t_f} w^T(\tau|t_f)Q^{-1}(\tau)w(\tau|t_f) + v^T(\tau|t_f)R^{-1}(\tau)v(\tau|t_f)\ d\tau$$

   where $v(\tau|t_f) = y(\tau) - C(\tau)x(\tau|t_f)$.

Thus, the goal is to use as little noise or initial condition as possible so as to *explain* the measured output $y(\tau)$.

Remarks:

- Set R^{-1} large when measurement is accurate, i.e. assume little measurement noise v.
- Set Q^{-1} large when process model is accurate, i.e. assume little process noise w.
- Set S^{-1} large when estimate of $\hat{x}(t_0|t_f) = 0$ is a confident guess of the initial state.

If a non-zero estimate (say x_0) for the initial state is desired, it is possible to modify the cost function so that:

$$\frac{1}{2}\hat{x}^T(t_0|t_f)S^{-1}\hat{x}(t_0|t_f) \Rightarrow \frac{1}{2}(\hat{x}(t_0|t_f) - x_0)^T S^{-1}(\hat{x}(t_0|t_f) - x_0).$$

It is clear that the optimal estimate at t_0 before any measurement is made is: $\hat{x}(t_0|t_0) = x_0$.

There are three types of Kalman filtering problem:

Filtering problem We are interested in solving for $\hat{x}(t|t_f)$ where $t < t_f$.

Prediction problem We are interested in solving $\hat{x}(t|t_f)$ where $t > t_f$.

Observer problem We are interested in $\hat{x}(t \mid t)$ where t is the current time. Thus, we need to solve $\hat{x}(t_f|t_f)$ and then determine how this varies with t_f. Conceptually, we need to resolve this problem when t_f changes. But, in practice, this is not necessary.

In the sequel, we shall drop the $B(t)u$ term in

$$\dot{x} = A(t)x + B(t)u$$

for convenience. It can easily be carried through.

7.3 Optimal filtering / observer Problem

The free variables in the optimal filtering / observer problem are:

- $\hat{x}(t_0|t_f)$, the estimate of the initial state and
- process noise $w(\tau)$.

Notice that once these have been chosen, $v(\tau|t_f) = y(\tau) - C(\tau)\hat{x}(\tau\ |t_f)$ is fixed.

Let the final time t_f be fixed. To avoid clutter, we shall omit the final time argument t_f. Thus, $\hat{x}(\tau)$ means $\hat{x}(\tau \mid x_f)$ etc.

The constraint of the system is:

$$\dot{\hat{x}}(\tau) = A\hat{x}(\tau) + B_w(\tau)w(\tau), \quad \tau \in [t_0, t_f].$$

We can convert this constrained optimal control problem into unconstrained optimization problem using the Lagrange multiplier $\lambda(\tau) \in \Re^n$.

The cost function for the unconstrained optimization problem is:

$$\bar{J} = \frac{1}{2}\hat{x}^T(t_0)S^{-1}\hat{x}(t_0) + \frac{1}{2}\int_{t_0}^{t_f} w^T Q^{-1}(\tau)w + v^T R^{-1}(\tau)v + 2\lambda^T(\tau)[\dot{\hat{x}} - (A(\tau)\hat{x} + B_w(\tau)w)]\ d\tau$$

Using integration by parts

$$\dot{\lambda x} = \lambda x - \dot{\lambda} x$$

and considering the variation of J with respect to variations in $\hat{x}(\tau)$, $w(\tau)$, $\hat{x}(t_f)$ and $\hat{x}(t_0)$:

$$\begin{aligned}\delta\bar{J} = &\int_{t_0}^{t_f} \left[w^T(\tau)Q^{-1}(\tau) - \lambda^T(\tau)B_w(\tau)\right]\delta w(\tau)\cdot d\tau \\ &+ \int_{t_0}^{t_f}\left[-v^T(\tau)R^{-1}(\tau)C^T(\tau) - \dot{\lambda}^T(\tau) - \lambda^T(\tau)A(\tau)\right]\delta\hat{x}(\tau)\cdot d\tau \\ &+ \lambda^T(t_f)\delta\hat{x}(t_f) + \left[\hat{x}(t_0)S^{-1} - \lambda^T(t_0)\right]\delta\hat{x}(t_0)\end{aligned}$$

and the variation with respect to $\lambda(\tau)$ is simply

$$+\int_{t_0}^{t_f} [\dot{\hat{x}} - (A(\tau)\hat{x} + B_w(\tau)w)]\delta\lambda(\tau)\cdot d\tau$$

Since $\delta x(\tau)$, $\delta w(\tau)$, $\delta\hat{x}(t_0)$, $\delta\hat{x}(t_f)$ as well as $\delta\lambda(\tau)$ can be arbitrary, the necessary condition for optimality is:

$$\begin{aligned}\delta w(\tau) &\to w(\tau) = Q(\tau)B_w^T(\tau)\lambda(\tau)\\ \delta\hat{x}(\tau) &\to \dot{\lambda}(\tau) = -A^T(\tau)\lambda(\tau) - \delta C^T(\tau)R^{-1}(\tau)[y(\tau) - C(\tau)\hat{x}(\tau)]\\ \lambda(\tau) &\to \dot{\hat{x}} = A(\tau)\hat{x} + B_w w\end{aligned}$$

giving the differential equations:

$$\begin{pmatrix}\dot{\hat{x}}\\ \dot{\lambda}\end{pmatrix} = \begin{pmatrix}A & B_w Q B_w^T\\ C^T R^{-1} C & -A^T\end{pmatrix}\begin{pmatrix}\hat{x}\\ \lambda\end{pmatrix} - \begin{pmatrix}0\\ C^T R^{-1} y(\tau)\end{pmatrix} \tag{7.4}$$

and the boundary conditions:

$$
\begin{aligned}
\delta\hat{x}(t_f) &\rightarrow \lambda(t_f) = 0 \\
\delta\hat{x}(t_0) &\rightarrow \lambda(t_0) = S^{-1}\hat{x}(t_0)
\end{aligned}
$$

This is a two point boundary value problem.

Notice that there is some similarity to the Hamilton-Jacobi-Bellman equation we saw for LQ control problem.

We solve this using a similar technique as for LQ to resolve the two point boundary value problem. Assume that

$$\hat{x}(t) = P(t)\lambda(t) + g(t) \tag{7.5}$$

Here $g(t)$ is used to account for the exogenous term $C^T(\tau)R^{-1}(\tau)y(\tau)$.

For (7.5) to be consistent with the HJB equations and boundary conditions, we need:

$$\dot{P} = A(\tau)P + PA^T(\tau) + B_w(\tau)Q(\tau)B_w^T(\tau) - PC^T(\tau)R^{-1}(\tau)C(\tau)P \tag{7.6}$$

$$\dot{g} = A(\tau)g - P(\tau)C^T(\tau)R^{-1}(\tau)[C(\tau)g - y(\tau)] \tag{7.7}$$

with initial condition $g(t_0) = 0$ and $P(t_0) = S$.

The first differential equation is the Filtering Riccati Differential Equation, which reduces to the CTRDE for LQ control if $A \rightarrow A^T$, $C \rightarrow B^T$ and $\tau \rightarrow (t_f - \tau)$.

Notice that all boundary conditions are assigned at $\tau = t_0$. Thus, (7.6)-(7.7) can be integrated forward in time.

Substituting (7.5) into (7.4),

$$\dot{\lambda} = -(A^T - C^T R^{-1} CP)\lambda + C^T R^{-1}[Cg(t) - y]; \quad \lambda(t_f) = 0. \tag{7.8}$$

$\lambda(t)$ can be integrated **backwards** in time, using $P(t)$ and $g(t)$ obtained from (7.6) and (7.7).

Once $\lambda(t)$ is obtained, the optimal estimate of the state at $\tau = t$ is then given by:

$$\hat{x}(t \mid t_f) = P(t|t_f)\lambda(t|t_f) + g(t|t_f).$$

Here we re-inserted the argument t_f to emphasize that the measurements $y(t)$ is available on the interval $[t_0, t_f]$.

Summary

Differential equations:

To be integrated forward in time

$$
\begin{aligned}
\dot{P} &= A(\tau)P + PA^T(\tau) + Q(\tau) - PC^T(\tau)R^{-1}(\tau)C(\tau)P \\
\dot{g} &= A(\tau)g - P(\tau)C^T(\tau)R^{-1}(\tau)[C(\tau)g - y(\tau)]
\end{aligned}
$$

To be integrated backward in time

$$\dot{\lambda} = -(A^T - C^T R^{-1} CP)\lambda - C^T R^{-1} y; \quad \lambda(t_f) = 0.$$

Boundary conditions:

- $g(t_0) = 0$ and $P(t_0) = S$, $\lambda(t_f) = 0$.

State estimate:

$$\hat{x}(t|t_f) = P(t|t_f)\lambda(t|t_f) + g(t|t_f).$$

7.4 Deterministic Optimal Observer

Let us denote $P(t \mid t_f)$ and $g(t \mid t_f)$ as the solutions to (7.6)-(7.7) with the available measurement up to t_f. To design an observer, we are interested in using available information up to current time to estimate the state at the same current time, i.e. we wish to find the solution $\hat{x}(t|t)$. This is done by taking $t_f = t$,

$$\hat{x}(t \mid t) = P(t \mid t)\lambda(t|t) + g(t|t) = g(t|t)$$

since $\lambda(t \mid t) = \lambda(t_f) = 0$.

Now, since the boundary conditions for (7.6)-(7.7) are given at t_0 (so they are integrated forward in time), so for any $\Delta > 0$,

$$P(t|t_f) = P(t|t_f + \Delta); \qquad g(t \mid t_f) = g(t|t_f + \Delta)$$

Hence, as t_f increases, the solution for $P(t|t_f)$ and $g(t|t_f)$ can be reused. In fact, the terminal time t_f does not enter into the solution of $P(\cdot)$ or $g(\cdot)$.

This means that the optimal observer is given by:

$$\begin{aligned}\frac{d}{dt}\hat{x}(t \mid t) = \dot{g}(t) &= A(t)g(t) - P(t)C^T(t)R^{-1}(t)[C(t)g(t) - y(t)] \\ &= A\hat{x}(t|t) - \underbrace{P(t)C^T(t)R^{-1}(t)}_{L(t)}[C(t)\hat{x}(t|t) - y(\tau)];\end{aligned} \tag{7.9}$$

with initial condition $\hat{x}(0 \mid 0) = x_0 = 0$, and

$$\dot{P} = A(t)P + PA^T(t) + B_w(t)Q(t)B_w^T(t) - PC^T(t)R^{-1}(t)C(t)P;$$

$P(t_0) = S$ and $L(t)$ is the observer gain.

- The Filtering Riccati Equation becomes the control Riccati Equation if we make the substitution:

$$A^T \to A, \quad C^T \to B, \qquad \tau \to (t_f - \tau)$$

  to obtain:

$$-\dot{P} = A^T(\tau)P + PA(\tau) + Q(\tau) - PB(\tau)R^{-1}(\tau)B^T(\tau)P$$

- The asymptotic solution $P(t \to \infty)$ for the Kalman Filtering problem is analogous to $P(t_0 \to -\infty)$ for the LQ control problem. Asymptotic properties of the Kalman filter can be obtained using this analogy.

- The ARE to be solved is obtained using the above substitutions:

$$AP_\infty + P_\infty A^T + Q - P_\infty C^T(t)R^{-1}CP_\infty = 0.$$

  The positive (semi-definite) solution P can be obtained via the eigenvectors of the Hamiltonian matrix in (7.4).

- In the time invariant case, $P(t \to \infty)$ converges if (A, C) is detectable.

- If $(A^T, Q^{\frac{1}{2}})$ is detectable, i.e. $(A, Q^{\frac{1}{2}T})$ is stabilizable, then

$$A_c = A - P_\infty C^T R^{-1} C$$

  is stable. Notice that although Q^{-1} is used in the performance function, the Riccati equation uses Q. So, it is possible to implement a Kalman filter with Q not invertible. In this case, this means that B_w does not have the maximum column rank and can be reduced.

7.5 Stochastic Kalman Filter

Consider a stochastic process:

$$\begin{aligned} \dot{x} &= Ax + w; \qquad x(t_0) = x_0 \\ y &= Cx + v \end{aligned} \tag{7.10}$$

Here $w(t)$ and $v(t)$ are random noises, and x_0 is a random variable: such that

$$\mathcal{E}[w(t)w^T(\tau)] = Q(t)\delta(t-\tau) \tag{7.11}$$

$$\mathcal{E}[v(t)v^T(\tau)] = R(t)\delta(t-\tau) \tag{7.12}$$

$$\mathcal{E}[v_i(t)w_j(\tau) = 0 \tag{7.13}$$

$$\mathcal{E}[x_0x_0^T] = S \tag{7.14}$$

where $\mathcal{E}(x)$ is the expected value (mean) of the random variable x. In other words,

- $w(t)$ is un-related to $w(\tau)$ whenever $\tau \neq t$.
- $v(t)$ is un-related to $v(\tau)$ whenever $\tau \neq t$.
- $v(t)$ is not related to $w(\tau)$ at all.
- $Q(t)$ is the expected size of $w(t)w^T(t)$.
- $R(t)$ is the expected size of $v(t)v^T(t)$.
- S is the expected size of $x_0x_0^T$.

Assume that $y(\tau)$ is measured for $\tau \in [t_0, t]$. If we apply the deterministic Kalman Filter to this stochastic process to obtain $\hat{x}(t|t)$, and let $\tilde{x}(t) = \hat{x}(t|t) - x(t)$. Then, it turns out that this estimate minimizes:

$$\mathcal{E}[\tilde{x}(t)\tilde{x}^T(t)]$$

Moreover, the solution to the Filtering Riccati Equation is

$$P(t) = \mathcal{E}[\tilde{x}(t)\tilde{x}^T(t)]$$

See Goodwin Chapter 22.10 for the derivation of this result.

7.5.1 Colored noise

If the expected noises are not white - i.e. $n(t)$ and $n(t+\tau)$ are related in some way (or $\mathcal{E}(n(t)n(t+\tau)) \neq 0$ for $t \neq \tau$), then one can augment the model with a noise model.

For example, if the noise w is low pass band limited to $[0, \omega_n]$, then we can assume that w is the output of an appropriate low pass filter with $[0, \omega_n]$ as the passband and a white noise input:

$$\begin{aligned} \dot{z} &= A_z z + B_z w_1 \\ w &= C_z z + D_z w_1 \end{aligned}$$

Here w is the band limited noise, but w_1 is white. In general, the noise shaping filter should be design so that it matches the frequency spectrum of the colored noise. This is so because the spectrum of a white noise is a constant so that the spectrum of the output of the shaping filter will be proportional to the frequency response of the filter itself.

Once the shaping filter has been designed, it can be incorporated into the process model so that the input to the system is now white. From the augmented model, the "standard" Kalman filter solution can be obtained.

7.6 Linear Quadratic Gaussian (LQG)

Consider the system driven by Gaussian noise:

$$\dot{x} = Ax + Bu + w$$
$$y = Cx + v$$

where

$$\mathcal{E}[w(t)w^T(\tau)] = Q_w(t)\delta(t-\tau) \tag{7.15}$$
$$\mathcal{E}[v(t)v^T(\tau)] = R_w(t)\delta(t-\tau) \tag{7.16}$$
$$\mathcal{E}[v_i(t)w_j(\tau) = 0 \tag{7.17}$$
$$\mathcal{E}[x_0 x_0^T] = S_w \tag{7.18}$$

Moreover, assume that the probability distributions of w and v are zero mean, and Gaussian.

Suppose that one wants to minimize:

$$J = \mathcal{E}\left\{\frac{1}{T}\int_{t_0}^{t_0+T} [x^T Q x + u^T R u]\ dt\right\}$$

where $T \to \infty$.

This is called the Linear Quadratic Gaussian (LQG) problem.

It turns out that the optimal control in this sense is given by:

$$u = -K\hat{x}$$

where

- K is the gain obtained by considering the LQR problem.
- $\hat{x}$ is the state estimate obtained from using the Kalman Filter.

From the separation theorem, we know that the observer estimated state feedback generates a stable controller such that the closed loop system has eigenvalues given by the eigenvalues of the observer error and the eigenvalues of the state feedback controller. The LQG result shows that Kalman Filter and LQ generates an Optimal Controller in the stochastic sense.

In fact, the LQR controller which assumes that the state $x(t)$ is directly available is optimal in the stochastic setting.

The LQG controller is optimal over ALL causal, linear and nonlinear controllers!

A drawback of LQG controller is that although it consists of a LQR and Kalman filter, both have good robustness properties, the LQG controller may not be robust. A design procedure called Loop Transfer Recovery (LTR) can sometimes be applied. The main idea of LTR is:

- Make Kalman filter look like state feedback
- This is achieved by making the Kalman filter rely on output injection rather than open loop estimation.
- This in turn is achieved by using large process noise (large Q) or small measurement noise (small R) models.
- For details, see (Anderson and Moore, 1990).