Kalman Filter

1960s. Rodolf E. Kalman published his famous discrete-time linear filtering problem in his paper
《A new approach to linear filtering and prediction problem》.

Rmk: A data processing technique for removing noise and restoring real data.

Linear-discrete SDE (controlled) → input control

(System) $x_k = F_k x_{k-1} + B_k u_k + w_k$ $x_k \in \mathbb{R}^n, F_k \in \mathbb{R}^{n\times n}, B_k \in \mathbb{R}^{n\times l}, u_k \in \mathbb{R}^l$

(Measurement) $z_k = H_k x_k + v_k$ $z_k \in \mathbb{R}^m, H_k \in \mathbb{R}^{m\times n}$

output

where $w_k \in N(0, Q_k)$, $v_k \in N(0, R_k)$, $Q_k \in \mathbb{R}^{n\times n}$, $R_k \in \mathbb{R}^{m\times m}$ are independent

notation:

- $\hat{x}_{k|k} = \mathbb{E}[x_k | \mathcal{F}_1, \cdots \mathcal{F}_k]$: estimate of state k at time k. ($\hat{x}_k := \hat{x}_{k|k}$)
- $\hat{x}_{k|k-1} = \mathbb{E}[x_k | \mathcal{F}_1, \cdots \mathcal{F}_{k-1}]$: prediction of state k at time $k-1$.
- $\hat{P}_{k|k}$: Posterior estimate error to measure accuracy

Idea: $\hat{x}_{k-1|k-1} \xrightarrow{\text{Prediction}} \hat{x}_{k|k-1} \xrightarrow[z_k]{\text{Correction}} \hat{x}_{k|k}$

State estimate : $\hat{x}_{k|k-1} = F_k \hat{x}_{k-1} + B_k u_k$

State error : $\tilde{x}_{k|k-1} = x_k - \hat{x}_{k|k-1}$ (Prediction)

Observation estimate: $\hat{z}_{k|k-1} = H_k \hat{x}_{k|k-1}$

Observation error: $\tilde{e}_k = z_k - \hat{z}_{k|k-1}$.

State error cov-matrix: $P_{k|k-1} = \text{cov}\{\tilde{x}_{k|k-1}\}$, $P_{k|k} = \text{cov}\{x_k - \hat{x}_{k|k}\}$

Observation error cov-matrix: $S_k = \text{cov}\{\tilde{e}_k\}$

Iteration : $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{e}_k$ K_k: Kalman gain $\in \mathbb{R}^{n\times m}$

Aim: Find K_k to minimize $\mathbb{E}[|x_k - \hat{x}_{k|k}|^2]$ (thus $tr(P_{k|k})$)

$$\begin{aligned}
P_{k|k-1} &= \text{cov}\{x_k - \hat{x}_{k|k-1}\} = \text{cov}\{F_k x_{k-1} + B_k u_k + w_k - F_k \hat{x}_{k-1} - B_k u_k\} \\
&= \text{cov}\{F_k(x_{k-1} - \hat{x}_{k-1}) + w_k\} = \text{cov}\{F_k(x_{k-1} - \hat{x}_{k-1})\} + \text{cov}\{w_k\} \\
&= F_k \,\text{cov}\{x_{k-1} - \hat{x}_{k-1}\} F_k^T + Q_k = F_k P_{k-1|k-1} F_k^T + Q_k.
\end{aligned}$$

Also $S_k = H_k P_{k|k-1} H_{k-1}^T + R_k$.

$= \text{cov}(H_k(x_k - \hat{x}_{k|k-1}) + v_k)$

$$P_{k-1|k-1} \longrightarrow P_{k|k-1} \xrightarrow{?} P_{k|k}$$

$$P_{k|k} = \operatorname{cov}\{\tilde{x}_{k|k}\} = \operatorname{cov}\{x_k - \hat{x}_{k|k}\} = \operatorname{cov}\{x_k - (\hat{x}_{k|k-1} + K_k \vec{e}_k)\}$$

$$= \operatorname{cov}\{x_k - (\hat{x}_{k|k-1} + K_k(z_k - \hat{z}_{k|k-1}))\}$$

$$= \operatorname{cov}\{x_k - (\hat{x}_{k|k-1} + K_k(H_k x_k + v_k - H_k \hat{x}_{k|k-1}))\}$$

$$= \operatorname{cov}\{(x_k - \hat{x}_{k|k-1}) - K_k H_k (x_k - \hat{x}_{k|k-1}) - K_k v_k\}$$

$$= \operatorname{cov}\{(I - K_k H_k)(x_k - \hat{x}_{k|k-1})\} + \operatorname{cov}\{K_k v_k\}$$

$$= (I - K_k H_k) P_{k|k-1} (I - K_k H_k)^T + K_k R_k K_k^T$$ expand, we get

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} - P_{k|k-1} H_k^T K_k^T + K_k S_k K_k^T$$

hence $\dfrac{\partial tr(P_{k|k})}{\partial K_k} = -2(H_k P_{k|k-1})^T + 2K_k S_k = 0$ ($S_k > 0$ as cov-matrix)

and $\dfrac{\partial^2 tr(P_{k|k})}{\partial K_k} = 2S_k > 0$ which implies when $K_k = P_{k|k-1} H_k^T S_k^{-1}$

$$K_k S_k K_k^T = P_{k|k-1} H_k^T K_k^T$$

Rmk: $\left(\dfrac{\partial tr(XA)}{\partial X}\right)_{ij} := \dfrac{\partial tr(XA)}{\partial x_{ij}} = \dfrac{\partial \sum_{k=1}^{n} \sum_{l=1}^{n} x_{kl} A_{lk}}{\partial x_{ij}} = A_{ji}$ $\quad \dfrac{\partial (tr(XA))}{\partial X} = A^T$

$\left(\dfrac{\partial tr(XAX^T)}{\partial X}\right)_{ij} := \dfrac{\partial tr(XAX^T)}{\partial x_{ij}} = \dfrac{\partial \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{q=1}^{n} x_{kl} A_{lq} x_{kq}}{\partial x_{ij}} = 2\sum_{q=1}^{n} A_{jq} x_{iq} = 2(XA)_{ij}$

$tr(P_{k|k})$ gets its infimum., and

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1}$$

Above all.

Step 1: $\hat{x}_{k-1|k-1} \Longrightarrow \hat{x}_{k|k-1}$ (Prediction with $\mathcal{F}_{k-1}$)

$$\begin{cases} \hat{x}_{k|k-1} = F_k \hat{x}_{k-1} + B_k u_k \\ P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \end{cases}$$

Step 2: $\hat{x}_{k|k-1} \Longrightarrow \hat{x}_{k|k}$ (Update with $\mathcal{F}_k$)

$$\begin{cases} \hat{e}_k = z_k - H_k \hat{x}_{k|k-1} \\ S_k = H_k P_{k|k-1} H_k^T + R_k > 0 \\ K_k = P_{k|k-1} H_k^T S_k^{-1} \end{cases}$$

$$\begin{cases} \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \hat{e}_k \\ P_{k|k} = (I - K_k H_k) P_{k|k-1} \end{cases}$$

The linear filtering Problem of continuous time.

Reference: « B. Oksendal »

(system) $dX_t = b(t, X_t)dt + \sigma(t, X_t)dU_t$

where $b: \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$, $\sigma: \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^{n\times p}$, U_t: p-dim B.M.

and X_0 is known and independent of U_t.

we assume observations $H_t \in \mathbb{R}^m$ are of the form.

$$H_t = c(t, X_t) + r(t, X_t)\cdot \tilde{W}_t \qquad \left(\tilde{W}_t = \frac{dV_t}{dt}\right).$$

where $c: \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^m$, $r: \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^{m\times r}$, $\tilde{W}_t$: r-dim white noise independent of U_t, X_0

we introduce $Z_t = \int_0^t H_s\, ds$, thereby

(Observations) $dZ_t = c(t, X_t)dt + r(t, X_t)dV_t$, $Z_0 = 0$

where V_t: r-dim B.M. independent of U_t and X_0

Rmk: If H_s is known for $0 \le s \le t$, then Z_s is also known for $0 \le s \le t$ and conversely. So no information is lost of gain by considering Z_t as our "observation" instead of H_t.

Qn: What is the "best estimate" $\hat{X}_t$ of X_t based on $\{Z_s, s \le t\}$

Aim: Find a SDE to characterize $\hat{X}_t$

define $\mathcal{G}_t = \sigma\{Z_s : s \le t\}$, $\hat{X}_t \subset \mathcal{G}_t \subset \mathcal{F}_t := \sigma\{U_s, V_s, X_0 : s \le t\}$

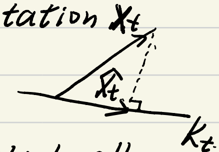
$(\Omega, \mathcal{F}, \mathbb{P})$ supports the (p+r)-dim B.M. (U_t, V_t)

$K_t := \{Y: \Omega \longrightarrow \mathbb{R}^n : Y \in L^2(\Omega, \mathbb{P})$ and $Y \subset \mathcal{G}_t\}$.

By saying $\hat{X}_t$ is the best estimate we mean that.

$$\int_\Omega |X_t - \hat{X}_t|^2 d\mathbb{P} = \mathbb{E}[|X_t - \hat{X}_t|^2] = \inf\{\mathbb{E}[|X_t - Y|^2] ; Y \in K_t\}.$$

In fact, recall the connection between conditional expectation and projection, we know



Lemma: $\hat{X}_t = \mathrm{Proj}_{K_t}(X_t) = \mathbb{E}[X_t | \mathcal{G}_t]$.

From now on we will concentrate on the linear case, which allows an explicit solution in terms of a SDE for $\hat{X}_t$ (the Kalman-Bucy filter)

Let's consider a simple, but motivating example.

Example. Suppose $X, W_1, W_2, \cdots$ are independent real r.v.s

$\mathbb{E}[X]=\mathbb{E}[W_j]=0$. $\mathbb{E}[X^2]=a^2$. $\mathbb{E}[W_j^2]=m^2$. for all j, put $Z_j=X+W_j$

Qn: What is the best linear estimate $\hat{X}$ of X based on $\{Z_j : j\le k\}$

Let $\mathcal{L}=\mathcal{L}(Z,k)=\operatorname{span}\{Z_1, Z_2, \cdots Z_k\}$. $\mathcal{F}_k:=\sigma\{Z_i : i\le k\}$

we want to find $\hat{X}_k=P_k(X):=\operatorname{Proj}_{\mathcal{L}(Z,k)}(X)=\mathbb{E}[X\mid\mathcal{F}_k]$

Core idea: Gram-Schmidt procedure.

Aim: obtain r.v.s $A_1, A_2, \cdots$ s.t.

(i) $\mathbb{E}[A_iA_j]=0$ if $i\ne j$) (ii) $\mathcal{L}(A,k)=\mathcal{L}(Z,k)$ for $\forall k$.

then
$$\hat{X}_k=\sum_{j=1}^{k}\frac{\mathbb{E}[XA_j]}{\mathbb{E}[A_j^2]}A_j \qquad \forall k=1,2\cdots$$

Let $A_1=Z_1$, for $j\ge 2$.

$$A_j:=Z_j-P_{j-1}(Z_j)=Z_j-P_{j-1}(X+W_j)=Z_j-\mathbb{E}[X+W_j\mid\mathcal{F}_{j-1}]=Z_j-\mathbb{E}[X\mid\mathcal{F}_{j-1}]$$

thus
$$A_j=Z_j-\hat{X}_{j-1} \qquad (1)$$

Also $\mathbb{E}[XA_j]=\mathbb{E}[X(Z_j-\hat{X}_{j-1})]=\mathbb{E}[X(X+W_j-\hat{X}_{j-1})]$ $\quad(\mathbb{E}[XW_j]=0)$

$=\mathbb{E}[X(X-\hat{X}_{j-1})]=\mathbb{E}[(X-\hat{X}_{j-1})^2]$ $\quad(\mathbb{E}[\hat{X}_{j-1}(X-\hat{X}_{j-1})]=0$

and $\mathbb{E}[A_j^2]=\mathbb{E}[(X+W_j-\hat{X}_{j-1})^2]=\mathbb{E}[(X-\hat{X}_{j-1})^2]+\mathbb{E}[W_j^2]=\mathbb{E}[(X-\hat{X}_{j-1})^2]+m^2$

hence.
$$\hat{X}_k=\hat{X}_{k-1}+\frac{\mathbb{E}[(X-\hat{X}_{k-1})^2]}{\mathbb{E}[(X-\hat{X}_{k-1})^2]+m^2}(Z_k-\hat{X}_{k-1}) \qquad \hat{X}_0=0$$

denote $\bar{Z}_k:=\frac{1}{k}\sum_{j=1}^{k}Z_j$ $\qquad \alpha_k:=\dfrac{a^2}{a^2+\frac{m^2}{k}}$, then $\hat{X}_k=\alpha_k\bar{Z}_k$

Obviously $\hat{X}_1=\dfrac{a^2}{a^2+m^2}Z_1$. By induction if $\hat{X}_{k-1}=\alpha_{k-1}\bar{Z}_{k-1}$

$$X-\hat{X}_{k-1}=X-\alpha_{k-1}\cdot\frac{1}{k-1}\sum_{j=1}^{k-1}(X+W_j)=(1-\alpha_{k-1})X-\frac{\alpha_{k-1}}{k-1}\sum_{j=1}^{k-1}W_j$$

$$\mathbb{E}[(X-\hat{X}_{k-1})^2]=(1-\alpha_{k-1})^2a^2+\frac{\alpha_{k-1}^2}{k-1}m^2=\frac{a^2m^2}{(k-1)a^2+m^2}$$

$$\hat{X}_k=\hat{X}_{k-1}+\frac{a^2}{ka^2+m^2}(Z_k-\hat{X}_{k-1})=\alpha_{k-1}\bar{Z}_{k-1}+\frac{a^2}{ka^2+m^2}[k\bar{Z}_k-(k-1)\bar{Z}_{k-1}-\alpha_{k-1}\bar{Z}_{k-1}]$$

$$=\alpha_k\bar{Z}_k$$

Above all, if $U_k := \alpha_k \bar{Z}_k$. (i). $U_k \in \mathcal{L}(Z,k)$, (ii) $X - U_k \perp \mathcal{L}(Z,k)$

in fact, we can check.

$$\begin{aligned}\mathbb{E}[(X-U_k)Z_i] &= \mathbb{E}[XZ_i] - \alpha_k \mathbb{E}[\bar{Z}_k Z_i] \\ &= \mathbb{E}[X(X+W_i)] - \alpha_k \frac{1}{k}\sum_j \mathbb{E}[Z_j Z_i] \\ &= a^2 - \frac{1}{k}\alpha_k (ka^2+m^2) = 0\end{aligned}$$

Corollary: For large k, $\hat{X}_k \approx Z_k$. (Law of large numbers)

In the linear filtering problem.

(linear system) $dX_t = F(t)X_t\,dt + C(t)\,dU_t$ $\quad F(t) \in \mathbb{R}^{n\times n}, C(t) \in \mathbb{R}^{n\times p}$

(linear observations) $dZ_t = G(t)X_t\,dt + D(t)\,dV_t$ $\quad G(t) \in \mathbb{R}^{m\times n}, D(t) \in \mathbb{R}^{m\times r}$

For simplicity, assume $m=n=p=r=1$

Assumption: (i) F, G, C, D are bounded on bounded intervals

(ii). $Z_0 = 0$, X_0 is normal distributed (independent of U_t, V_t)

(iii). $D(t)$ is bounded away from 0 on bounded intervals.

Lemma. Gaussian r.v.s are closed in the L^2 limitation sense,

with Picard interation we can know

$$M_t = \begin{bmatrix} X_t \\ Z_t \end{bmatrix} \text{ is a Gaussian process.}$$

Step1: Denote $\mathcal{L} = \mathcal{L}(Z,t) = \overline{\{c_0 + c_1 Z_{s_1} + \cdots + c_k Z_{s_k} : s_j \le t, c_j \in \mathbb{R}\}}^{L^2(\mathbb{P})}$

Then $\hat{X}_t = P_K(X_t) = \mathbb{E}[X_t | \mathcal{G}_t] = P_{\mathcal{L}}(X_t)$.

Step2: Replace Z_t by the innovation process N_t:

$$N_t := Z_t - \int_0^t P_{\mathcal{L}}(G(s)X_s)\,ds = Z_t - \int_0^t G(s) P_{\mathcal{L}}(X_s)\,ds$$

i.e. $dN_t = G(t)(X_t - \hat{X}_t)dt + D(t)\,dV_t$

Then (1). N_t has orthogonal increment i.e.

$$\mathbb{E}[(N_{t_1} - N_{s_1})(N_{t_2} - N_{s_2})] = 0 \text{ if } [s_1,t_1) \cap [s_2,t_2] = \phi.$$

(ii) $\mathcal{L}(N,t) = \mathcal{L}(Z,t)$, so $\hat{X}_t = P_{\mathcal{L}(N,t)}(X_t)$, furthermore

$$\mathcal{L}(Z,T) = \{c_0 + \int_0^T f(t)\,dZ_t : f \in L^2[0,T],\ c_0 \in \mathbb{R}\}. \quad (2)$$

(iii). N_t is a Gaussian process.

Rmk: In fact, the Lebesgue integral of a Gaussian process is Gaussian under some mild measurability and integrability condition

Step 3. If we put $dR_t = \frac{1}{D(t)} dN_t$ $(D(t) > 0)$
we can observe that:

(i) R_t has continuous paths
(ii) R_t has orthogonal increments (since N_t has)
(iii) R_t is Gaussian (since N_t is)
(iv) $\mathbb{E}[R_t] = 0$ and $\mathbb{E}[R_t R_s] = \min(s,t)$ (Ito formula to $d(R_t^2)$)

Properties (i), (iii) and (iv) imply R_t is a B.M.. Moreover
$\mathcal{L}(N,t) = \mathcal{L}(R,t)$, with (2) and R_t is a B.M., we obtain

$$\hat{X}_t = \mathbb{E}[X_t] + \int_0^t \frac{\partial}{\partial s} \mathbb{E}[X_t R_s] \, dR_s \qquad (3)$$

Step 4. Find an expression of X_t where $dX_t = F(t) X_t dt + C(t) dU_t$

$$X_t = \exp\left\{\int_0^t F(s) ds\right\} X_0 + \int_0^t \exp\left\{\int_s^t F(u) du\right\} C(s) dU_s \qquad (4)$$

Step 5. Substitute (4) into $\mathbb{E}[X_t R_s]$ and use (3) to obtain a SDE.
From (4) we know $\mathbb{E}[X_t] = \exp\{\int_0^t F(s) ds\} \mathbb{E}[X_0]$.

$$d\hat{X}_t = \frac{\partial}{\partial s} \mathbb{E}[X_t R_s]\big|_{s=t} dR_t + \left(\int_0^t \frac{\partial^2}{\partial t \partial s} \mathbb{E}[X_t R_s] dR_s\right) dt + F(t) \mathbb{E}[X_t]$$

After some calculus, let $S(t) := \mathbb{E}[(X_t - \hat{X}_t)^2]$.
then $S(t)$ satisfies Riccati equation (one-dim)

$$\frac{dS}{dt} = 2F(t) S(t) - \frac{G^2(t)}{D^2(t)} S^2(t) + C^2(t), \quad S(0) = \mathbb{E}[(X_0 - \mathbb{E}[X_0])^2].$$

Thm (The 1-dim Kalman-Bucy filter)
The solution $\hat{X}_t = \mathbb{E}[X_t | \mathcal{G}_t]$ of the 1-dim linear filtering problem
(linear system) $dX_t = F(t) X_t dt + C(t) dU_t$; $F(t), C(t) \in \mathbb{R}$
(linear observations) $dZ_t = G(t) X_t dt + D(t) dV_t$; $G(t), D(t) \in \mathbb{R}$.
satisfies the SDE

$$d\hat{X}_t = \left(F(t) - \frac{G^2 S(t)}{D^2(t)}\right) \hat{X}_t dt + \frac{G(t) S(t)}{D^2(t)} dZ_t \; ; \quad \hat{X}_0 = \mathbb{E}[X_0]$$

Rmk: The Riccati equation is solvable under some suitable condition.

Example 1 (Noisy observations of a constant process).

(system). $dX_t = 0$, i.e. $X_t = X_0$; $\mathbb{E}[X_0] = \hat{X}_0$, $\mathbb{E}[X_0^2] = a^2$

(observations) $dZ_t = X_t dt + m dV_t$; $Z_0 = 0$

corresponding to $H_t = \frac{dZ}{dt} = X_t + m\frac{dV_t}{dt}$ where $\frac{dV_t}{dt}$ is white noise.

Riccati equation: $S(t) = \mathbb{E}[(X_t - \hat{X}_t)^2]$

$$\frac{dS}{dt} = -\frac{1}{m^2}S^2 \,, \quad S(0) = a^2 \Longrightarrow S(t) = \frac{a^2 m^2}{m^2 + a^2 t} \,;$$

$$d\hat{X}_t = -\frac{a^2}{m^2+a^2 t}\hat{X}_t dt + \frac{a^2}{m^2+a^2 t} dZ_t$$

$$\Longrightarrow \quad \hat{X}_t = \frac{m^2}{m^2+a^2 t}\hat{X}_0 + \frac{a^2}{m^2+a^2 t} Z_t .$$

More examples and the theorem for multi-dim Kalman-Bucy Filter, see «Oksendal»

Thm. (The Muti-Dim Kalman-Bucy Filter).

The solution $\hat{X}_t = \mathbb{E}[X_t \mid \mathcal{G}_t]$ of the multi-dim linear filtering problem

(linear system) $dX_t = F(t)X_t dt + C(t)dU_t$; $F(t) \in \mathbb{R}^{n\times n}$, $C(t) \in \mathbb{R}^{n\times p}$

(linear observations) $dZ_t = G(t)X_t dt + D(t)dV_t$; $G(t) \in \mathbb{R}^{m\times n}$, $D(t) \in \mathbb{R}^{m\times r}$

satisfies the SDE:

$$d\hat{X}_t = (F - SG^T(DD^T)^{-1}G)\hat{X}_t dt + SG^T(DD^T)^{-1}dZ_t \,; \quad \hat{X}_0 = \mathbb{E}[X_0]$$

where $S(t) := \mathbb{E}[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T] \in \mathbb{R}^{n\times n}$ satisfies the matrix Riccati equation

$$\frac{dS}{dt} = FS + SF^T - SG^T(DD^T)^{-1}GS + CC^T$$

$$S(0) = \mathbb{E}[(X_0 - \mathbb{E}[X_0])(X_0 - \mathbb{E}[X_0])^T]$$

The condition on $D(t) \in \mathbb{R}^{m\times r}$ is now is invertible for all t and $(D(t)D(t)^T)^{-1}$ is bounded on any bdd interval.