

1-dim linear filtering problem:

(linear dyn.)

(linear observations.)

$$\begin{cases} dx_t = F(t) x_t dt + C(t) dw_t, \\ dz_t = G(t) x_t dt + D(t) dv_t, \end{cases}$$

Assumptions:

→ let's assume $z_0 = 0$, and $x_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and independent of w_t, v_t .

and F_t, C_t, G_t, D_t are all bounded on finite intervals and

and D_t is bounded away from zero on finite intervals ($D_t^T D_t > 0$).

The best MSE estimator $\hat{x}_t$ of x_t that is $G_t = \{z_s: 0 \leq s \leq t\}$ -meas
is just the conditional

$$\hat{x}_t = E[x_t | G_t] = P_{L(G_t)}(x_t) = E[x_t] + \int_0^t \frac{\partial}{\partial s} E[x_t R_s] dR_s$$

where

$$dR_t = \frac{1}{D_t} dN_t = \frac{1}{D_t} [dz_t - G_t \hat{x}_t dt] = \frac{G_t}{D_t} (x_t - \hat{x}_t) dt + dv_t$$

The 1-dim MSE optimal estimator is the Kalman-Bucy filter:

$$d\hat{x}_t = \left[F_t - \frac{G_t^2}{D_t^2} S_t \right] \hat{x}_t dt + \frac{G_t}{D_t^2} S_t dz_t, \quad \hat{x}_0 = E[x_0]$$

where $S_t \triangleq E[\tilde{x}_t^2] = E[(x_t - \hat{x}_t)^2]$ satisfies the Riccati equation

$$\frac{dS_t}{dt} = 2F_t S_t - \frac{G_t^2}{D_t^2} S_t^2 + C_t^2, \quad S_0 = E[(x_0 - E(x_0))^2]$$

How did we derive the Riccati eq. of $\dot{S}_t$?

$$\rightarrow S_t = E[\tilde{x}_t^2] = E[x_t^2] - E[\hat{x}_t^2] \quad \text{by } \begin{array}{c} x_t \\ \nearrow \\ \hat{x}_t \end{array} \tilde{x}_t$$

$$= E[x_t^2] - E[(\hat{x}_t - E[\hat{x}_t])^2] - E[\hat{x}_t]^2$$

Now, by $E[\hat{x}_t] = E[x_t]$:

$$S_t = E[x_t^2] - E[(\hat{x}_t - E[x_t])^2] - E[x_t]^2$$

by Itô Isometry

$$= E[x_t^2] - \int_0^t f^2(s, t) ds - E[x_t]^2$$

$$E[x_t^2] = E[\varphi^2(0, t) x_0^2] + E\left[\left(\int_0^t \varphi(s, t) C_s dw_s\right)^2\right] \quad \text{as } x_0 \perp w_s$$

by Itô Isometry

$$= \varphi^2(0, t) E[x_0^2] + \int_0^t \varphi^2(s, t) C_s^2 ds$$

So

$$S_t = \varphi^2(0, t) [E[x_0^2] - E[x_0]^2] + \int_0^t [\varphi^2(s, t) C_s^2 - f^2(s, t)] ds$$

Recall

$$\frac{d}{dt} \left[\int_0^t g(s, t) ds \right] = g(t, t) + \int_0^t \frac{\partial}{\partial t} g(s, t) ds$$

$$\Rightarrow \frac{dS_t}{dt} = 2 F_t \varphi(0, t) [E(x_0^2) - E(x_0)^2] + \varphi^2(t, t) C_t^2 - f^2(t, t)$$

$$+ \int_0^t \left[2 F_s \varphi(s, t) C_s^2 - 2 f(s, t) \cdot \frac{\partial f}{\partial t}(s, t) \right] ds$$

$$+ \int_0^t \dots$$

$$\text{but } f(s,t) = \frac{G_s}{D_s} \varphi(s,t) S_s \rightarrow \begin{cases} f(t,t) = \frac{G_t}{D_t} S_t \text{ and} \\ \frac{\partial}{\partial t} f(s,t) = \frac{G_s}{D_s} F_t \varphi(s,t) S_s \\ = F_t \cdot f(s,t) \end{cases}$$

$$\text{and } C_0(t) = \mathbb{E}[X_t] = \varphi(0,t) \mathbb{E}[X_0] \rightarrow C_0'(t) = F_t \varphi(0,t) \mathbb{E}[X_0] = F_t \cdot C_0(t)$$

$$S_0: \frac{dS_t}{dt} = 2 F_t \varphi(0,t) \underbrace{[\mathbb{E}(X_0^2) - \mathbb{E}(X_0)^2]} + C_t^2 - \frac{G_t^2}{D_t^2} S_t^2 + 2 F_t \int_0^t \underbrace{[\varphi(s,t) C_s^2 - f(s,t)]} ds$$

$$\text{by } S_t = 2 F_t S_t + C_t^2 - \frac{G_t^2}{D_t^2} S_t^2$$

$$\text{Riccati Equation: } \boxed{\dot{S}_t = 2 F_t S_t + C_t^2 - \frac{G_t^2}{D_t^2} S_t^2}$$

Exercise: Find $\mathbb{E}[\hat{X}_t]$ and $\mathbb{E}[\hat{X}_t^2]$ explicitly!
(Hint: Use def. of S_t and $\hat{X}_t$.)

Examples: noisy observations of a constant process

(Constant dyn) $dx_t = 0$, and X_0 is Gaussian $\begin{cases} \mathbb{E}[X_0] = \mu_0 \\ \mathbb{E}[X_0^2] = \sigma^2 \end{cases}$

$\therefore 1, \dots, m, dV_t, Z_0 = 0$ "white noise!"

(linear observ.) $dz_t = x_t dt + m \tilde{W}_t$

↳ now, this corresponds to $H_t = \frac{dz_t}{dt} = x_t + m \tilde{W}_t$
 but we said that we don't have a well-defined mathematical model of white noise which we wanna think as derivative of a wiener process " $\frac{dW_t}{dt}$ " ! that is why we worked with z_t instead

s.t. $z_t = \int_0^t H_s ds \rightsquigarrow$ essentially a sum over such "white noise" which is well-defined!

→ What is the Kalman-Bucy filter?

solve the Riccati equation for $S_t = E[|x_t - \hat{x}_t|^2]$:

$$\frac{dS_t}{dt} = -\frac{1}{m^2} S_t^2, \quad S(0) = a^2$$

$$\Rightarrow S_t = \frac{a^2 m^2}{m^2 + a^2 t}, \quad t \geq 0$$

so the estimate $\hat{x}_t$:

$$d\hat{x}_t = \underbrace{-\frac{a^2}{m^2 + a^2 t} \hat{x}_t}_{F_t} dt + \frac{a^2}{m^2 + a^2 t} dz_t, \quad \hat{x}_0 = E[x_0]$$

Ex: Use the integrating factor $\exp(-\int_0^t F_s ds)$ to get:

$$\hat{x}_t = \frac{m^2}{m^2 + a^2 t} \hat{x}_0 + \frac{a^2}{m^2 + a^2 t} z_t, \quad t \geq 0, \quad \hat{x}_0 = E[x_0]$$

Example 6.2.10 (Noisy observations of a Brownian motion).

If we modify the preceding example slightly, so that

$$\begin{aligned} \text{(system)} \quad & dX_t = cdU_t; \quad E[X_0] = 0, \quad E[X_0^2] = a^2, \quad c \text{ constant} \\ \text{(observations)} \quad & dZ_t = X_t dt + mdV_t, \end{aligned}$$

the Riccati equation becomes

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

or

$$\frac{m^2 dS}{m^2 c^2 - S^2} = dt, \quad (S \neq mc) .$$

This gives

$$\left| \frac{mc + s}{mc - s} \right| = K \exp \left(\frac{2ct}{m} \right); \quad K = \left| \frac{mc + a^2}{mc - a^2} \right| .$$

Or

$$S(t) = \begin{cases} mc \frac{K \cdot \exp(\frac{2ct}{m}) - 1}{K \cdot \exp(\frac{2ct}{m}) + 1}; & \text{if } S(0) < mc \\ mc \text{ (constant)} & \text{if } S(0) = mc \\ mc \frac{K \cdot \exp(\frac{2ct}{m}) + 1}{K \cdot \exp(\frac{2ct}{m}) - 1} & \text{if } S(0) > mc . \end{cases}$$

Thus in all cases the mean square error $S(t)$ tends to mc as $t \rightarrow \infty$.

For simplicity let us put $a = 0$, $m = c = 1$. Then

$$S(t) = \frac{\exp(2t) - 1}{\exp(2t) + 1} = \tanh(t) .$$

The equation for $\hat{X}_t$ is

$$d\hat{X}_t = -\tanh(t)\hat{X}_t dt + \tanh(t)dZ_t, \quad \hat{X}_0 = 0$$

or

$$d(\cosh(t)\hat{X}_t) = \sinh(t)dZ_t .$$

So

$$\hat{X}_t = \frac{1}{\cosh(t)} \int_0^t \sinh(s) dZ_s .$$

If we return to the interpretation of Z_t :

$$Z_t = \int_0^t H_s ds ,$$

where H_s are the “original” observations (see (6.1.4)), we can write

$$\hat{X}_t = \frac{1}{\cosh(t)} \int_0^t \sinh(s) H_s ds , \quad (6.2.31)$$

so $\hat{X}_t$ is approximately (for large t) a weighted average of the observations H_s , with increasing emphasis on observations as time increases.

Example 6.2.11 (Estimation of a parameter).

Suppose we want to estimate the value of a (constant) parameter θ , based on observations Z_t satisfying the model

$$dZ_t = \theta M(t)dt + N(t)dB_t ,$$

where $M(t), N(t)$ are known functions. In this case the stochastic differential equation for θ is of course

$$d\theta = 0 ,$$

so the Riccati equation for $S(t) = E[(\theta - \hat{\theta}_t)^2]$ is

$$\frac{dS}{dt} = - \left(\frac{M(t)S(t)}{N(t)} \right)^2$$

which gives

$$S(t) = \left(S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds \right)^{-1}$$

and the Kalman-Bucy filter is

$$d\hat{\theta}_t = \frac{M(t)S(t)}{N(t)^2}(dZ_t - M(t)\hat{\theta}_t dt) .$$

This can be written

$$\left(S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds\right) d\hat{\theta}_t + M(t)^2 N(t)^{-2} \hat{\theta}_t dt = M(t) N(t)^{-2} dZ_t .$$

We recognize the left hand side as

$$d\left(\left(S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds\right) \hat{\theta}_t\right)$$

so we obtain

$$\hat{\theta}_t = \frac{\hat{\theta}_0 S_0^{-1} + \int_0^t M(s) N(s)^{-2} dZ_s}{S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds} .$$

This estimate coincides with the maximum likelihood estimate in classical estimation theory if $S_0^{-1} = 0$. See Liptser and Shiryaev (1978).

For more information about estimates of drift parameters in diffusions and generalizations, see for example Aase (1982), Brown and Hewitt (1975) and Taraskin (1974).

Theorem 6.3.1 (The Multi-Dimensional Kalman-Bucy Filter).

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

$$\begin{aligned} \text{(linear system)} \quad dX_t &= F(t)X_t dt + C(t)dU_t; \\ &F(t) \in \mathbf{R}^{n \times n}, \quad C(t) \in \mathbf{R}^{n \times p} \end{aligned} \quad (6.3.1)$$

$$\begin{aligned} \text{(linear observations)} \quad dZ_t &= G(t)X_t dt + D(t)dV_t; \\ &G(t) \in \mathbf{R}^{m \times n}, \quad D(t) \in \mathbf{R}^{m \times r} \end{aligned} \quad (6.3.2)$$

satisfies the stochastic differential equation

$$d\hat{X}_t = (F - SG^T(DD^T)^{-1}G)\hat{X}_t dt + SG^T(DD^T)^{-1}dZ_t; \quad \hat{X}_0 = E[X_0] \quad (6.3.3)$$

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

$$\begin{aligned} \frac{dS}{dt} &= FS + SF^T - SG^T(DD^T)^{-1}GS + CC^T ; \\ S(0) &= E[(X_0 - E[X_0])(X_0 - E[X_0])^T] . \end{aligned} \quad (6.3.4)$$

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

A similar solution can be found for the more general situation

$$\text{(system)} \quad dX_t = [F_0(t) + F_1(t)X_t + F_2(t)Z_t]dt + C(t)dU_t \quad (6.3.5)$$

$$\text{(observations)} \quad dZ_t = [G_0(t) + G_1(t)X_t + G_2(t)Z_t]dt + D(t)dV_t , \quad (6.3.6)$$

where $X_t \in \mathbf{R}^n$, $Z_t \in \mathbf{R}^m$ and $B_t = (U_t, V_t)$ is $n + m$ -dimensional Brownian motion, with appropriate dimensions on the matrix coefficients. See Bensoussan (1992) and Kallianpur (1980), who also treat the non-linear case. An account of non-linear filtering theory is also given in Pardoux (1979) and Davis (1984).

For the solution of linear filtering problems governed by more general processes than Brownian motion (processes with orthogonal increments) see Davis (1977).

For various applications of filtering theory see Bucy and Joseph (1968), Jazwinski (1970), Gelb (1974), Maybeck (1979) and the references in these books.

The Prediction Problem :

Recall that in the filtering problem we found the best MSE estimator $\hat{X}_t$ of X_t given $G_t = \{z_s : 0 \leq s \leq t\}$.

Now, the Prediction problem is to find the best MSE estimator N_T of X_T given $G_t = \{z_s : 0 \leq s \leq t\}$ with $T > t$.

with exactly the same argument as filtering problem:

$$N_T = P_{\mathcal{L}^2(\mathcal{G}_t)}(X_T) = \mathbb{E}[X_T | \mathcal{G}_t]$$

but $dx_t = F_t X_t dt + C_t dw_t$ so

$$X_T = \varphi(t, T) X_t + \int_t^T \varphi(s, T) C_s dw_s$$

so

$$N_T = \mathbb{E}[\varphi(t, T) X_t | \mathcal{G}_t] + \underbrace{\mathbb{E}\left[\int_t^T \varphi(s, T) C_s dw_s \mid \mathcal{G}_t\right]}_{= 0}$$

$$= \varphi(t, T) \mathbb{E}[X_t | \mathcal{G}_t] + \underbrace{\mathbb{E}\left[\int_t^T \varphi(s, T) C_s dw_s\right]}_{= 0}$$

$$= \varphi(t, T) \hat{X}_t$$

$$= \exp\left(\int_t^T F_s ds\right) \cdot \hat{X}_t$$

Exercise: what is $\mathbb{E}[N_T]$ and $\mathbb{E}[N_T^2]$ explicitly?

The Smoothing / Interpolation Problem:

In contrast to the prediction problem, given $\mathcal{G}_t = \{z_s: 0 \leq s \leq t\}$

the smoothing problem ask for finding the best MSE estimation

m_s of X_s for any $\underline{s < t}$.

Again a similar argument shows that

$$M_s = P_{L^2(\mathcal{G}_t)}(X_s) = E[X_s | \mathcal{G}_t]$$

- Find an explicit formula for M_s for $s \leq t = \frac{dM_s}{ds} = ?!$

→ note that $M_t = P_{L^2(\mathcal{G}_t)}(X_t) = \hat{X}_t$.

$$\begin{aligned} \rightarrow M_s = E[X_s | \mathcal{G}_t] &= E[\hat{X}_s | \mathcal{G}_t] + E[X_s - \hat{X}_s | \mathcal{G}_t] \\ &= \hat{X}_s + E[\tilde{X}_s | \mathcal{G}_t] \quad \text{as } \hat{X}_s \in \mathcal{G}_s \subset \mathcal{G}_t \end{aligned}$$

Recall $\hat{X}_t = \hat{X}_s + \int_s^t \frac{\partial}{\partial r} E[X_t R_r] dR_r \leftarrow s \leq r \leq t$ but what is $E[\tilde{X}_s | \mathcal{G}_t]$?!

$$M_s = P_{L(R,t)}(X_s) = E[X_s] + \int_0^t \frac{\partial}{\partial r} E[X_s R_r] dR_r \rightarrow \text{by a similar argument as Kalman filter as } R_t \text{ has orth. increments}$$

$$= E[X_s] + \int_0^s \frac{\partial}{\partial r} E[X_s R_r] dR_r + \int_s^t \frac{\partial}{\partial r} E[X_s R_r] dR_r$$

(note the difference with $\frac{\partial}{\partial r} E[X_t R_r]$) $= \hat{X}_s + \int_s^t \frac{\partial}{\partial r} E[X_s R_r] dR_r$

$$\text{and } dR_t = \frac{G_t}{\sigma_t^2} \tilde{X}_t dt + dV_t$$

$$(s \leq r \leq t) \Rightarrow E[X_s R_r] = E\left[X_s \int_0^r \frac{G_z}{\sigma_z^2} \tilde{X}_z dz\right] + E[X_s V_r]$$

!! $(X_s \perp V_r)$

$$\Rightarrow \frac{\partial}{\partial r} E[x_s R_r] = \frac{G_r}{D_r} E[x_s \tilde{x}_r]$$

$$\begin{aligned} d\tilde{x}_t &= dx_t - d\hat{x}_t = F_t x_t dt + C_t dw_t - \hat{F}_t \hat{x}_t dt - \frac{G_t}{D_t^2} S_t (dz_t - G_t \hat{x}_t dt) \\ &= F_t \tilde{x}_t dt + C_t dw_t - \underbrace{\frac{G_t}{D_t^2} S_t (dz_t - G_t \hat{x}_t dt)}_{= dV_t \text{ innovation process}} \\ &= \left(F_t - \frac{G_t^2}{D_t^2} S_t \right) \tilde{x}_t dt + C_t dw_t - \frac{G_t}{D_t} S_t dV_t \end{aligned}$$

Kalman filter error dynamics :

$$d\tilde{x}_t = \left(F_t - \frac{G_t^2}{D_t^2} S_t \right) \tilde{x}_t dt + C_t dw_t - \frac{G_t}{D_t} S_t dV_t$$

transition matrix

$$\Rightarrow \tilde{x}_r = \tilde{\varphi}(s, r) \tilde{x}_s + \underbrace{\int_s^r \varphi(s, \tau) \left[C_\tau dw_\tau - \frac{G_\tau}{D_\tau} S_\tau dV_\tau \right]}_{\perp x_s \text{ (future noise)}}$$

$$\Rightarrow E[x_s \tilde{x}_r] = \tilde{\varphi}(s, r) E[x_s \tilde{x}_s]$$

$$\Rightarrow \frac{\partial}{\partial r} E[x_s R_r] = \frac{G_r}{D_r} \tilde{\varphi}(s, r) \cdot E[\tilde{x}_s^2] = \frac{G_r}{D_r} \tilde{\varphi}(s, r) S_s$$

$$\begin{aligned} \text{so, } m_s &= \hat{x}_s + \int_s^t \frac{G_r}{D_r} \tilde{\varphi}(s, r) S_s dR_r \\ &= \hat{x}_s + S_s \cdot \underbrace{\int_s^t \frac{G_r}{D_r} \tilde{\varphi}(s, r) dR_r}_{\triangleq \Lambda_s} \end{aligned}$$

$\Rightarrow$ Think of $\tilde{\varphi}(s, r) = \exp\left(\int_s^r \left(F_\tau - \frac{G_\tau^2}{D_\tau^2} S_\tau\right) d\tau\right)$ as the transition matrix

$(r > s)$ from " r " to " s " and $\frac{\partial \tilde{\varphi}(s, r)}{\partial s} = -\left(F_s - \frac{G_s^2}{D_s^2} S_s\right) \tilde{\varphi}(s, r)$ then

$$\Rightarrow \begin{cases} d\Lambda_s = -\left(F_s - \frac{G_s^2}{D_s^2} S_s\right) \Lambda_s ds - \frac{G_s}{D_s} dR_s & (s < t) \\ \Lambda_t = 0 \end{cases}$$

Now. $m_s = \hat{x}_s + S_s \cdot \Lambda_s \quad \leadsto \quad \Lambda_s = S_s^{-1} (m_s - \hat{x}_s)$

$$\Rightarrow dm_s = d\hat{x}_s + \dot{S}_s \Lambda_s ds + S_s \cdot d\Lambda_s$$

(a magical elimination of dR_s term)

$$= F_s \hat{x}_s ds + \frac{G_s}{D_s^2} S_s \underbrace{(dz_s - G_s \hat{x}_s ds)}_{dN_s = D_s dR_s} + \dot{S}_s \Lambda_s ds$$

$$- \left(F_s S_s - \frac{G_s^2}{D_s^2} S_s^2\right) \Lambda_s ds - \frac{G_s}{D_s} S_s dR_s$$

$$= F_s \hat{x}_s ds + \left(\dot{S}_s - F_s S_s + \frac{G_s^2}{D_s^2} S_s^2\right) \Lambda_s ds$$

(by Riccati equation)

$$= F_s \hat{x}_s ds + \left(F_s S_s + C_s^2\right) \Lambda_s ds$$

(by def of Λ_s)

$$= F_s \hat{x}_s ds + \left(F_s + C_s^2 S_s^{-1}\right) (m_s - \hat{x}_s) ds$$

$$= F_s m_s ds + C_s^2 S_s^{-1} (m_s - \hat{x}_s) ds$$

In summary: the smoothing solution m_s given G_t solve the backward ODE

$$\boxed{\begin{aligned} \frac{dm_s}{ds} &= F_s m_s + C_s^2 S_s^{-1} (m_s - \hat{x}_s), \quad s \leq t \\ m_t &= \hat{x}_t \end{aligned}}$$

This also proves that $m_s = E[x_s | G_t]$ is smooth for $s < t$.

In matrix case :

$$\left\{ \begin{array}{l} \frac{dm_s}{ds} = F_s m_s + C_s C_s^T S_s^{-1} (m_s - \hat{x}_s) , \quad s \leq t \\ m_t = \hat{x}_t \end{array} \right.$$