

7. KALMAN FILTER (generalization)

7.1. $\mathbf{E}X_0 \neq 0$, $\text{cov}(X_0, Y_0) \neq 0$.

For $\mathbf{E}X_0 = 0$ and $\text{cov}(X_0, Y_0) = 0$, the Kalman filter was derived in previous lecture. Now, we give the Kalman filter without these restrictions.

Set $\hat{X}_0 = \mathbf{E}(X_0|Y_0)$ and

$$\begin{aligned} X_t^\circ &= X_t - \hat{X}_0 \exp\left(\int_0^t a(s)ds\right) \\ Y_t^\circ &= Y_t - \hat{X}_0 \int_0^t A(s) \exp\left(\int_0^s a(u)du\right) ds. \end{aligned}$$

Since $X_0^\circ = X_0 - \mathbf{E}(X_0|Y_0)$ and $Y_0^\circ = Y_0$, we have $\mathbf{E}X_0^\circ = 0$, $\text{cov}(X_0^\circ, Y_0^\circ) = 0$. So,

$$\begin{aligned} dX_t^\circ &= a(t)X_t^\circ dt + b(t)dW_t' \\ dY_t^\circ &= A(t)X_t^\circ dt + B(t)dW_t'' \end{aligned}$$

and thereby

$$\begin{aligned} d\hat{X}_t^\circ &= a(t)\hat{X}_t^\circ dt + \frac{P^\circ(t)A(t)}{B^2(t)}(dY_t^\circ - A(t)\hat{X}_t^\circ dt) \\ \dot{P}^\circ(t) &= 2a(t)P^\circ(t) + b^2(t) - \left(\frac{P^\circ(t)A(t)}{B(t)}\right)^2 \end{aligned}$$

subject to $\hat{X}_0^\circ = 0$, $P(0) = \mathbf{E}(X_0 - \hat{X}_0^\circ)^2$.

It is clear that for any $t \geq 0$,

$$\{Y_s^\circ, s \leq t\} = \{Y_s, s \leq t\}.$$

Consequently, for any $t \geq 0$

$$\mathbf{E}(X_t^\circ | (Y^\circ)_t) = \mathbf{E}(X_t^\circ | Y_0^t).$$

We use this equality to derive equations for $\hat{X}_t = \mathbf{E}(X_t | Y_0^t)$ and $P(t) = \mathbf{E}(X_t - \hat{X}_t)^2$.

Since (Home work)

$$X_t - \hat{X}_t = X_t^\circ - \hat{X}_t^\circ \tag{7.1}$$

we have

$$P(t) \equiv P^\circ(t),$$

i.e.

$$\dot{P}(t) = 2a(t)P(t) + b^2(t) - \left(\frac{P(t)A(t)}{B(t)}\right)^2$$

subject to $P(0) = \mathbf{E}(X_0 - \mathbf{E}(X_0|Y_0))^2$.

Further,

$$\begin{aligned}\widehat{X}_t &= \mathbf{E}(X_t|Y_0^t) = \mathbf{E}(X_t|(Y_0^y)^\circ) \\ &= \widehat{X}_t^\circ + \widehat{X}_0 \exp\left(\int_0^t a(s)ds\right),\end{aligned}$$

that is (Home work)

$$d\widehat{X}_t = a(t)\widehat{X}_t dt + \frac{P(t)A(t)}{B^2(t)}(dY_t - A(t)\widehat{X}_t dt) \quad (7.2)$$

with $\widehat{X}_0 = \mathbf{E}(X_0|Y_0)$.

7.2. Generalized Kalman filter.

Consider now more complicated "signal-observation" model (recall that X_t is the signal and Y_t is the observation):

$$\begin{aligned}dX_t &= [a(t)X_t + c_0(t) + c_1(t)u(t, Y_0^t)]dt + b_1(t)dW_t' + b_2(t)dW_t'' \\ dY_t &= [A(t)X_t + C_0(t) + C_1(t)v(t, Y_0^t)]dt + B(t)dW_t''.\end{aligned} \quad (7.3)$$

Here W_t' , W_t'' are independent Wiener processes, $a(t)$, $A(t)$, $b_i(t)$, $B(t)$, $c_j(t)$, $C_j(t)$, $i = 1, 2$; $j = 0, 1$ are continuous deterministic functions, and $u(t, Y_0^t)$, $v(t, Y_0^t)$ are functions of arguments t, Y_s , $0 \leq s \leq t$ such that equations (7.3) obey the unique solution and

$$\int_0^T \mathbf{E}[u^2(t, Y_0^t) + v^2(t, Y_0^t)]dt < \infty.$$

The function $B^2(t)$ is assumed to be uniformly positive ($B^2(t) \geq \gamma > 0$). Here, we assume also that (X_0, Y_0) is Gaussian vector.

Theorem. (Without proof) *The optimal (in the mean square sense) filtering estimate \widehat{X}_t and the filtering error $P(t) = \mathbf{E}(X_t - \widehat{X}_t)^2$ are defined by the generalized Kalman filter:*

$$\begin{aligned}d\widehat{X}_t &= [a(t)\widehat{X}_t + c_0(t) + c_1(t)u(t, Y_0^t)]dt \\ &\quad + \frac{b_2(t)B(t) + P(t)A(t)}{B^2(t)}\left(dY_t - [A(t)\widehat{X}_t + C_0(t) + C_1(t)v(t, Y_0^t)]dt\right) \\ \dot{P}(t) &= 2a(t)P(t) + b_1^2(t) + b_2^2(t) - \frac{(b_2(t)B(t) + P(t)A(t))^2}{B^2(t)}\end{aligned} \quad (7.4)$$

subject to the initial conditions $\widehat{X}_0 = \mathbf{E}(X_0|Y_0)$ and $P(0) = \mathbf{E}(X_0 - \widehat{X}_0)^2$.

7.3. Innovation Wiener process.

Set

$$\bar{W}_t = \int_0^t \frac{1}{B(s)} \left(dY_s - [A(s)\hat{X}_s + C_0(s) + C_1(s)v(s, Y_0^s)] ds \right)$$

Theorem. $(\bar{W}_t)_{t \geq 0}$ is Wiener process.

Proof. Due to the second Itô equation in (7.3),

$$\bar{W}_t = W_t'' + \int_0^t \frac{A(s)}{B(s)} [X_s - \hat{X}_s] ds.$$

The direct computations show that $\bar{W}_0 = 0$ and $\mathbf{E}\bar{W}_t = 0$. Let us check that $\mathbf{E}\bar{W}_{t'}\bar{W}_{t''} = t' \wedge t''$. For $t'' > t'$,

$$\begin{aligned} \mathbf{E}\bar{W}_{t'}\bar{W}_{t''} &= \mathbf{E}(\bar{W}_{t'})^2 + \mathbf{E}\bar{W}_{t'}[\bar{W}_{t''} - \bar{W}_{t'}] \\ &= \mathbf{E}(\bar{W}_{t'})^2 + \int_{t'}^{t''} \mathbf{E}\bar{W}_{t'} \frac{A(s)}{B(s)} [X_s - \hat{X}_s] ds \\ &= \mathbf{E}(\bar{W}_{t'})^2 + \int_0^t \mathbf{E}\bar{W}_{t'} \frac{A(s)}{B(s)} \mathbf{E}[X_s - \hat{X}_s | Y_0^s] ds \\ &= \mathbf{E}(\bar{W}_{t'})^2 \\ &= t'. \end{aligned}$$

For $t'' < t'$, in the same way, we get $\mathbf{E}\bar{W}_{t'}\bar{W}_{t''} = t''$. Particularly, $\mathbf{E}\bar{W}_t^2 = t$. Thus, it remains to check only that \bar{W}_t is Gaussian process. To this end, use the Itô formula. Let $f(t)$ be piece wise constant deterministic function. Then the Itô integral $\int_0^t f(s) d\bar{W}_s$ forms a linear combinations of increments of the $(\bar{W}_t)_{t \geq 0}$ process. Further, we use the following obvious fact: if for every f and $t > 0$ the random variable $\int_0^t f(s) d\bar{W}_s$ is Gaussian, then any linear combination $\sum_i c_i \bar{W}_{t_i}$ is Gaussian random variable what is nothing but the definition of a Gaussian process.

Introduce the random process $Z_t = \int_0^t f(s) d\bar{W}_s$ and, for $\lambda \in \mathbb{R}$, define the random process $e^{i\lambda Z_t}$, where $i = \sqrt{-1}$. By the Itô formula, applying to $e^{i\lambda Z_t}$, we find

$$e^{i\lambda Z_t} = 1 + \int_0^t i\lambda f(s) e^{i\lambda Z_s} d\bar{W}_s - \frac{\lambda^2}{2} \int_0^t f^2(s) e^{i\lambda Z_s} ds. \quad (7.5)$$

Taking the expectation from both parts of this equality we arrive at the equation for the characteristic function $F_t(\lambda) = \mathbf{E}e^{i\lambda Z_t}$ the Fourier transform of the random variable Z_t :

$$F_t(\lambda)_t = 1 - \int_0^t \frac{\lambda^2 f^2(s)}{2} F_s(\lambda) ds$$

(recall that $\mathbf{E} \int_0^t f(s) e^{i\lambda Z_s} d\bar{W}_s = 0$). Therefore

$$F_t(\lambda) = e^{-\frac{\lambda^2}{2} \int_0^t f^2(s) ds},$$

i.e. the characteristic function corresponds to the Gaussian distribution with parameters $(0, \int_0^t f^2(s) ds)$. \square

7.4. Conditionally Gaussian filter.

We consider now a nonlinear "signal-observation" model for which there exists a filter similar to the generalized Kalman filter.

$$\begin{aligned} dX_t &= [a_0(t, Y_0^t) + a_1(t, Y_0^t)X_t]dt + b_1(t, Y_0^t)dW_t' + b_2(t, Y_0^t)dW_t'' \\ dY_t &= [A_0(t, Y_0^t) + A_1(t, Y_0^t)X_t]dt + B(t, Y_0^t)dW_t'', \end{aligned} \quad (7.6)$$

where W_t', W_t'' are independent Wiener processes, functions $a_0(t, Y_0^t), \dots, B(t, Y_0^t)$ are such that equations (7.6) obey the unique solution and $a_1(t, Y_0^t)$ and $A_1(t, Y_0^t)$ are bounded and $B^2(t, Y_0^t) \geq c > 0$.

Theorem. (Without proof.) *Let the conditional distribution $P(X_0 \leq x|Y_0)$ be Gaussian (P -a.s.) with parameters $\hat{X}_0 = \mathbf{E}(X_0|Y_0)$ and $P(0) = \mathbf{E}(X_0 - \hat{X}_0)^2|Y_0$. Then for any $t > 0$ the conditional distribution $P(X_t \leq x|Y_0^t)$ is Gaussian (P -a.s.), its parameter $\hat{X}_t = \mathbf{E}(X_t|Y_0^t)$ and $P(t) = \mathbf{E}(X_t - \hat{X}_t)^2|Y_0^t$ are defined by the filtering equations*

$$\begin{aligned} d\hat{X}_t &= [a_0(t, Y_0^t) + a_1(t, Y_0^t)\hat{X}_t]dt + \frac{b_2(t, Y_0^t)B(t, Y_0^t) + P(t)A(t)}{B^2(t, Y_0^t)} \\ &\quad \times \left(dY_t - [A_0(t, Y_0^t) + A_1(t, Y_0^t)\hat{X}_t]dt \right) \\ \dot{P}(t) &= 2a_1(t, Y_0^t)P(t) + b_1^2(t, Y_0^t) + b_2^2(t, Y_0^t) \\ &\quad - \frac{(b_2(t, Y_0^t)B(t, Y_0^t) + P(t)(A_1(t, Y_0^t)))^2}{B^2(t)} \end{aligned} \quad (7.7)$$

subject to \hat{X}_0 and $P(0)$.

At all

$$\bar{W}_t = \int_0^t \frac{1}{B(s, Y_0^s)} \left(dY_s - [A_0(s, Y_0^s) + A_1(s, Y_0^s)\hat{X}_s]ds \right)$$

is Wiener process (innovation).

Remark 1. The nonlinear filter, defined in (7.7) is called the conditionally Gaussian filter. Particularly, the extended Kalman filter is conditionally Gaussian.

Remark 2. For the conditionally Gaussian filter the mean square filtering error is equal to $\mathbf{E}P(t)$.

7.5. Extended Kalman filter.

The conditionally Gaussian model is specified by drifts linear in X_t for both equations: ‘ $dX_t = \dots$ ’ and ‘ $dY_t = \dots$ ’. Now consider a model with nonlinear in X_t drifts:

$$\begin{aligned} dX_t &= a(t, X_t)dt + b_1(t)dW_t' + b_2(t)dW_t'' \\ dY_t &= A(t, X_t)dt + B(t)dW_t''. \end{aligned} \tag{7.8}$$

For this model, the optimal in the mean square sense filter yields essentially more complicated structure than any of the above-mentioned filters. Therefore, it makes sense to find a filter which is not optimal, but not “far” from optimal, having a simple structure similar to that the conditionally Gaussian filter has. From common sense of view such a filter might be created, if the function $a(t, x)$ and $A(t, x)$ are closed to linear ones.

To this end, assume that the intensity $B^2(t)$ of the noise $B(t)dW_t''$ is too small and so a good filtering accuracy can be expected. Denote by \widehat{X}_t a filtering estimate. Assuming the filtering error $X_t - \widehat{X}_t$ is small enough and the functions $a(t, x)$ and $A(t, x)$ are twice continuously differentiable in x and their derivatives $a'_x(t, x), a''_{xx}(t, x), A'_x(t, x), A''_{xx}(t, x)$ are bounded one can use an approximations

$$\begin{aligned} a(t, X_t) &\approx a(t, \widehat{X}_t) + a'_x(t, \widehat{X}_t)[X_t - \widehat{X}_t] \\ A(t, X_t) &\approx A(t, \widehat{X}_t) + A'_x(t, \widehat{X}_t)[X_t - \widehat{X}_t]. \end{aligned} \tag{7.9}$$

Therefore, one can use the structure of the conditionally Gaussian filter corresponding to

$$\begin{aligned} dX_t' &= \left(a(t, \widehat{X}_t') + a'_x(t, \widehat{X}_t')[X_t' - \widehat{X}_t'] \right) dt + b_1(t)dW_t' + b_2(t)dW_t'' \\ dY_t &= \left(A(t, \widehat{X}_t') + A'_x(t, \widehat{X}_t')[X_t' - \widehat{X}_t'] \right) dt + B(t)dW_t'' \end{aligned}$$

and obtain the extended Kalman filter

$$\begin{aligned} d\widehat{X}_t &= a(t, \widehat{X}_t)dt + \frac{b_2(t)B(t) + P(t)A'_x(t, \widehat{X}_t)}{B^2(t)} \\ &\quad \times \left(dY_t - A(t, \widehat{X}_t)dt \right) \\ \dot{P}(t) &= 2a'_x(t, \widehat{X}_t)P(t) + b_1^2(t) + b_2^2(t) \\ &\quad - \frac{(b_2(t)B(t) + P(t)(A'_x(t, \widehat{X}_t)))^2}{B^2(t)}. \end{aligned}$$

Not always this filter is good enough. Below we give a simplest analysis for the extended Kalman filter.

Assume

$$\begin{aligned}dX_t &= a(X_t)dt + dW'_t \\dY_t &= X_t + dW''_t\end{aligned}$$

and the function $a(x)$ is Lipschitz continuous $|a(x) - a(y)| \leq \ell|x - y|$ with known parameter ℓ . The extended Kalman filter is defined as:

$$d\widehat{X}_t = a(\widehat{X}_t)dt + P(t)[dY_t - \widehat{X}_t dt],$$

where the Kalman gain $P(t)$ has to be defined by given the above Ricatti equation. Here, we apply another approach being effective, if $\ell \leq a_x \geq c > 0$. Let $P(t)$ be some non negative function which has to be chosen. To this end, let us introduce a random process

$$\alpha_t = \begin{cases} \frac{a(X_t) - a(\widehat{X}_t)}{X_t - \widehat{X}_t} & X_t \neq \widehat{X}_t, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\Delta_t = X_t - \widehat{X}_t$ satisfies

$$d\Delta_t = \alpha_t \Delta_t dt + dW'_t - P(t)dW''_t - P(t)\Delta_t dt.$$

Applying the Itô formula to Δ_t^2 we find

$$\begin{aligned}\Delta_t^2 &= \Delta_0^2 + 2 \int_0^t [\alpha_s - P(s)] \Delta_s^2 ds + t + \int_0^t P^2(s) ds \\ &\quad + 2 \int_0^t \Delta_s dW'_s - 2 \int_0^t \Delta_s dW''_s.\end{aligned}\tag{7.10}$$

Set $\Gamma_t = \mathbf{E}\Delta_t^2$. The expectation taken from both sides of (7.10) provides

$$\Gamma_t = \Gamma_0 + 2 \int_0^t \mathbf{E}[\alpha_s - P(s)] \Delta_s^2 ds + t + \int_0^t P^2(s) ds$$

and therefore

$$\begin{aligned}\dot{\Gamma}_t &= 2\mathbf{E}[\alpha_t - P(t)]\Delta_t^2 + 1 + P^2(t) \\ &\leq 2[\ell - P(t)]\Gamma_t + 1 + P^2(t).\end{aligned}$$

Since $P(t)$ is an arbitrary function, one can to minimize in $P(t)$ the right side of the above differential inequality, that is

$$\begin{aligned}\dot{\Gamma}_t &\leq \min_{P(t)} \left\{ 2[\ell - P(t)]\Gamma_t + 1 + P^2(t) \right\} \\ &= 2\ell\Gamma_t + 1 - \Gamma_t^2.\end{aligned}\tag{7.11}$$

Parallel to (7.11) introduce the Riccati equation

$$\dot{\Gamma}_t^\circ = 2\ell\Gamma_t^\circ + 1 - (\Gamma_t^\circ)^2.$$

subject to $\Gamma_0^\circ = \Gamma_0$.

It holds (Home work)

$$\Gamma_t \leq \Gamma_t^\circ, \forall t > 0. \quad (7.12)$$

So, $P(t) \equiv \Gamma^\circ(t)$.

Home work.

1. Prove (7.1).
2. Prove (7.2).
3. Prove (7.12). **Hint:** Show that $\Delta_t = \Gamma_t^\circ - \Gamma_t$ obeys the property:

$$\dot{\Delta}_t \geq \{2\ell - (\Gamma_t^\circ + \Gamma_t)\}\Delta_t.$$