6. KALMAN-BUCY FILTER

6.1. **Motivation and preliminary.**

As was shown in Lecture 2, the optimal control is a function of all coordinates of controlled process. Very often, it is not impossible to observe a controlled process or part of its component. For instance, an information on a controlled trajectory is interrupted by a noise. In this case the optimal (or reasonable) control can not be constructed as a functional of controlled trajectories.

It is very natural to estimate a controlled trajectory via existing observations and then to create a control as a functional of these estimates. The Kalman-Bucy filter is one of filters which enables us to have on line estimating, or tracking, of unobservable signals.

A standard filtering problem deals with two random processes, say X_t and Y_t, where X_t is an unobservable signal and Y_t is an observable one. The filtering problem consists in the estimating of X_t via $\{Y_s, s \le t\}$, that is any filtering estimate $\widehat{X}_t$ is a functional of $Y_0^t = \{Y_s, 0 \le s \le t\}$. The more popular optimal in the mean square sense filtering estimate is the conditional expectation

$$\widehat{X}_t = \boldsymbol{E}(X_t \big| Y_0^t)$$

minimizing the mean square error

$$P(t) = \boldsymbol{E}(X_t - \widehat{X}_t)^2.$$

The conditional expectation $\boldsymbol{E}(X_t \big| Y_0^t)$, as a function of Y_0^t, may have too complicated structure. Therefore, especially in engineering applications, it makes sense to use "a linear function of Y_0^t" minimizing the mean square error $P(t)$ in a class of linear filtering estimates. The optimal linear estimate $\widehat{X}_t$ is the orthogonal projection on a linear space generated by Y_0^t[1]. According to the optimal property of the orthogonal projection it is named " the conditional expectation in the wide sense". For Gaussian processes the orthogonal projection coincides with the conditional expectation.

6.2. **Kalman-Busy model.**

Let the signal X_t and the observation Y_t are defined by the Itô linear equations with respect to independent Wiener processes W_t' and W_t''

$$\begin{aligned} dX_t &= \quad a(t)X_t dt + b(t)dW_t' \\ dY_t &= \; A(t)X_t dt + B(t)dW_t'' \end{aligned} \tag{6.1}$$

[1]this linear space is generated by any linear combination of Y_{t_j} "+ constant" and their closure in the mean square sense

subject to Gaussian initial conditions X_0 and Y_0 independent of W'_t and W''_t. The deterministic (known) functions $a(t)$, $A(t)$, $b(t)$, and $B(t)$ are assumed to be bounded and and piece-wise continuous for $t \geq 0$.

The main assumption here is

$$\inf_{t\geq 0} B^2(t) \geq c \text{ (for some } c > 0). \tag{6.2}$$

6.2.1. **Description of the Kalman-Bucy filter.**

The filtering estimate $\widehat{X}_t$ and the mean square filtering error $P(t)$ are defined by differential equations (Itô and Riccati)

$$\begin{aligned} d\widehat{X}_t &= a(t)\widehat{X}_t dt + \frac{P(t)A(t)}{B^2(t)}(dY_t - A(t)\widehat{X}_t dt) \\ \dot{P}(t) &= 2a(t)P(t) + b^2(t) - \frac{P^2(t)A^2(t)}{B^2(t)} \end{aligned} \tag{6.3}$$

subject to the initial conditions:

$$\begin{aligned} \widehat{X}_0 &= \boldsymbol{E}X_0 + \frac{\operatorname{cov}(X_0, Y_0)}{\operatorname{cov}(Y_0, Y_0)}(Y_0 - \boldsymbol{E}Y_0) \\ P(0) &= \operatorname{cov}(X_0, X_0) - \frac{\operatorname{cov}^2(X_0, Y_0)}{\operatorname{cov}(Y_0, Y_0)}. \end{aligned} \tag{6.4}$$

6.2.2. **Derivation of filtering equations.**

(6.4) is Home work.

Since here we deal with Gaussian random processes the conditional expectation coincides with the orthogonal projection which is a linear function of observations. Let us denote $\widehat{X}_t = \mathcal{L}(Y_0^t)$. To describe a structure of $\mathcal{L}(Y_0^t)$, introduce partitions of the interval $[0, t]$: $0 = t_0^n < t_1^n < ... < t_n^n = t$ so that $\lim_{n\to\infty} \max_j [t_{j+1}^n - t_j^n] = 0$ and find the optimal in the mean square filtering estimate of X_t given $\{Y_{t_j^n}, j = 0, ..., n\}$. Since $\{Y_{t_j^n}, j = 0, ..., n\}$ and $\{Y_0, Y_{t_j^n} - Y_{t_{j-1}^n}, j = 1, ..., n\}$ contain the same "information", we have

$$\mathcal{L}^n(Y_0^t) = \boldsymbol{E}(X_t | Y_0, Y_{t_j^n} - Y_{t_{j-1}^n}, j = 1, ...n)$$

and, whereas $\mathcal{L}^n(Y_0^t)$ is a linear function of arguments $Y_0, Y_{t_j^n} - Y_{t_{j-1}^n}, j = 1, ...n$, we have

$$\mathcal{L}^n(Y_0^t) = G_0^n(t) + G_1^n(t)Y_0 + \sum_j G^n(t, t_j^n)[Y_{t_j^n} - Y_{t_{j-1}^n}],$$

where $G_0^n(t), G_1^n(t)$ and $G^n(t, t_j^n)$'s are deterministic function.

We start with a particular case $\boldsymbol{E}X_0 = 0$ and $\mathrm{cov}(X_0, Y_0) = 0$. Notice that (6.4) provides $G_0^n(0) \equiv 0$, $G_1^n(0) \equiv 0$. Hence, we have

$$\mathcal{L}^n(Y_0^t) = \sum_j G^n(t, t_j^n)[Y_{t_j^n} - Y_{t_{j-1}^n}] = \int_0^t G^n(t,s)dY_s,$$

where $G^n(t,s) = G(t, t_j^n)$, $t_j^n \le s < t_{j+1}^n$.

For further convenience, we choose partitions

$$\{t_1^n, \dots, t_n^n\} \subseteq \{t_1^{n+1}, \dots, t_{n+1}^{n+1}\}, \ n \ge 1,$$

so that

$$\{Y_0, Y_{t_j^n} - Y_{t_{j-1}^n}, j = 1, ..., n\} \uparrow \{Y_0 = 0, Y_s, s \le t\}.$$

Then, by the well known property of the conditional expectation

$$\underset{n\to\infty}{\mathrm{l.i.m.}}\, \boldsymbol{E}\Big(X_t \Big| Y_0, Y_{t_j^n} - Y_{t_{j-1}^n}, j = 1, ..., n\Big) = \boldsymbol{E}\Big(X_t \Big| Y_0, Y_s, s \le t\Big),$$

that is

$$\boldsymbol{E}(X_t | Y_0^t) = \underset{n\to\infty}{\mathrm{l.i.m.}}\, \mathcal{L}^n(Y_0^t).$$

Our next task is to show that there exists a function $G(t,s)$ such that

$$\boldsymbol{E}(X_t | Y_0^t) = \int_0^t G(t,s)dY_s. \tag{6.5}$$

By the Cauchy criteria for the convergence in the mean square sense we have

$$\underset{n\to\infty}{\mathrm{l.i.m.}}\, \mathcal{L}^n(Y_0^t) \text{ exists} \Longleftrightarrow \lim_{n,m\to 0} \boldsymbol{E}\Big(\mathcal{L}^n(Y_0^t) - \mathcal{L}^m(Y_0^t)\Big)^2 = 0.$$

On the other hand, since $\mathcal{L}^n(Y_0^t)$ is the linear function of Y_0^t, we have

$$\mathcal{L}^n(Y_0^t) - \mathcal{L}^m(Y_0^t) := \int_0^t [G^n(t,s) - G^m(t,s)]dY_s$$

and

$$\lim_{n,m\to 0} \boldsymbol{E}\Big(\int_0^t [G^n(t,s) - G^m(t,s)]dY_s\Big)^2 = 0. \tag{6.6}$$

Therefore, taking into account the independence of (X_t) and (W_t'') and denoting by $K(s', s'') = \boldsymbol{E}(X_{s'}X_{s''})$ (see Home work 2.), write

$$
\begin{aligned}
&\boldsymbol{E}\Big(\int_0^t [G^n(t,s)-G^m(t,s)]dY_s\Big)^2 \\
= \;& \boldsymbol{E}\Big(\int_0^t [G^n(t,s)-G^m(t,s)][A(s)X_s ds + B(s)dW''_s]\Big)^2 \\
= \;& \boldsymbol{E}\Big(\int_0^t [G^n(t,s)-G^m(t,s)]A(s)X_s ds\Big)^2 \\
& + \boldsymbol{E}\Big(\int_0^t [G^n(t,s)-G^m(t,s)]B(s)dW''_s\Big)^2 \\
= \;& \int_0^t\int_0^t [G^n(t,s')-G^m(t,s')][G^n(t,s'')-G^m(t,s'')]K(s',s'')ds'ds'' \\
& + \int_0^t [G^n(t,s)-G^m(t,s)]^2 B^2(s)ds.
\end{aligned}
$$

Both terms in the right side of this equality are nonnegative and therefore limits, as $n,m\to\infty$, for both are zeros. Moreover, since $B^2(s)\ge c>0$, we obtain

$$
\lim_{n,m\to\infty}\int_0^t [G^n(t,s)-G^m(t,s)]^2 ds = 0.
$$

The latter, by the Cauchy criteria provides (6.5) and, moreover,

$$
\begin{aligned}
\underset{n\to\infty}{\text{l.i.m.}}\int_0^t G^n(t,s)A(s)X_s ds &= \int_0^t G(t,s)A(s)X_s ds \\
\underset{n\to\infty}{\text{l.i.m.}}\int_0^t G^n(t,s)B(s)dW''_s &= \int_0^t G(t,s)B(s)dW''_s.
\end{aligned}
$$

Consequently, (6.5) holds true.

6.2.3. **Wiener-Hopf equation.**

Now, we show that $G(t,s)$ solves the Wiener-Hopf integral equation: for every fixed t and $s\le t$

$$
A(s)K(t,s) = \int_0^t G(t,u)A(s)K(s,u)A(u)du + G(t,s)B^2(s),
$$

where $K(t,s)$ is the correlation function of the process (X_t) and $A(s)$, $B(s)$ are function involved in (6.1).

Proof: Since $\int_0^t G(t,s)dY_s$ is the orthogonal projection generated by Y_0^t, the random variable $X_t-\int_0^t G(t,s)dY_s$ is orthogonal to any linear function of Y_0^t. Then, for

$\int_0^t f(t,s)dY_s$ with deterministic (bounded) function $f(t,s)$, it holds

$$\boldsymbol{E}\Big(X_t - \int_0^t G(t,s)dY_s\Big)\int_0^t f(t,s)dY_s = 0.$$

The latter equality implies

$$\boldsymbol{E}\Big(X_t\int_0^t f(t,s)dY_s\Big) = \boldsymbol{E}\Big(\int_0^t G(t,s)dY_s\int_0^t f(t,s)dY_s\Big).$$

Write, keeping in mind the independence of (X_t) and (W''_t),

$$\begin{aligned}
\boldsymbol{E}X_t\int_0^t f(t,s)dY_s &= \boldsymbol{E}X_t\Big(\int_0^t f(t,s)A(s)X_s ds + \int_0^t f(t,s)B(s)dW''_s\Big)\\
&= \boldsymbol{E}X_t\int_0^t f(t,s)A(s)X_s ds\\
&= \int_0^t f(t,s)A(s)\boldsymbol{E}(X_tX_s)ds\\
&= \int_0^t f(t,s)A(s)K(t,s)ds \qquad (6.7)
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{E}\int_0^t G(t,s)dY_s\int_0^t f(t,s)dY_s &= \boldsymbol{E}\Big(\int_0^t G(t,s)A(s)X_s ds + \int_0^t G(t,s)B(s)dW''_s\Big)\\
&\quad\times\Big(\int_0^t f(t,s)A(s)X_s ds + \int_0^t f(t,s)B(s)dW''_s\Big)\\
&= \boldsymbol{E}\int_0^t G(t,s)A(s)X_s ds\int_0^t f(t,s)A(s)X_s ds\\
&\quad+\boldsymbol{E}\int_0^t f(t,s)B(s)dW''_s\int_0^t G(t,s)B(s)dW''_s\\
&= \int_0^t\int_0^t f(t,s)G(t,u)A(u)K(s,u)A(s)duds\\
&\quad+\int_0^t f(t,s)G(t,s)B^2(s)ds. \qquad (6.8)
\end{aligned}$$

(6.7) and (6.8) provide the equality

$$\begin{aligned}
\int_0^t f(t,s)\Big\{A(s)K(t,s) &- \int_0^t G(t,u)A(u)K(s,u)A(s)du\\
&- f(t,s)G(t,s)B^2(s)\Big\}ds = 0.
\end{aligned}$$

So, by an arbitrariness of $f(t,s)$, we claim

$$A(s)K(t,s) = \int_0^t G(t,u)A(s)K(s,u)A(u)du + G(t,s)B^2(s).$$

Uniqueness of solution of the Wiener-Hopf equation.

Assume $G_i(t,s)$, $i=1,2$ are solutions. Set $\Delta(t,s) = G_1(t,s) - G_2(t,s)$. Then

$$\int_0^t \Delta(t,s)A(s)K(s,u)A(u)ds + \Delta(t,u)B^2(u) = 0$$

Multiplying both sides of this equality by $\Delta(t,s)$ and integrating with respect to ds we get

$$\int_0^t \int_0^t \Delta(t,s)A(s)K(s,u)A(u)\Delta(t,u)duds + \int_0^t \Delta^2(t,u)B^2(u)du = 0.$$

Both parts on the left hand side of above equality equal zero. Since $B^2(u)$ is strictly positive $\Delta^2(t,u) = 0$, $u \le t$.

Existence of solution of the Wiener-Hopf equation.

1. Show first that

$$G(t,t) = \frac{P(t)A(t)}{B^2(t)}, \tag{6.9}$$

$$\begin{aligned}
G(t,t)B^2(t) &= K(t,t)A(t) - \int_0^t G(t,u)A(t)K(t,u)A(u)du \\
&= \boldsymbol{E}X_t^2A(t) - \int_0^t G(t,u)A(t)\boldsymbol{E}(X_tX_u)A(u)du \\
&= \boldsymbol{E}\Big(\Big[X_t - \int_0^t G(t,u)A(u)X_u du\Big]X_tA(t)\Big) \\
&= \boldsymbol{E}\Big(\Big[X_t - \int_0^t G(t,u)A(u)X_u du - \int_0^t G(t,u)B(u)dW''_u\Big]X_tA(t)\Big) \\
&= \boldsymbol{E}\Big(\Big[X_t - \int_0^t G(t,u)dY_u\Big]X_tA(t)\Big) \\
&= \boldsymbol{E}\Big(\Big[X_t - \widehat{X}_t\Big]X_tA(t) = \Big[X_t - \widehat{X}_t\Big]\Big[X_t - \widehat{X}_t\Big]A(t) \\
&\quad + \Big[X_t - \widehat{X}_t\Big]\widehat{X}_tA(t)\Big) \\
&= P(t)A(t) + \boldsymbol{E}\Big(\Big[X_t - \widehat{X}_t\Big]\widehat{X}_tA(t)\Big)
\end{aligned}$$

and since $\boldsymbol{E}\Big(\Big[X_t - \widehat{X}_t\Big]\widehat{X}_t\Big) = 0$ we obtain $G(t,t)B^2(t) = P(t)A(t)$ and (6.9).

2. For $s < t$, we will find s solution in a form

$$G(t,s) = \varphi_s^t G(s,s) \tag{6.10}$$

with smooth in t function φ_s^t, $\varphi_t^t = 1$.

Notice that $K(t,s)$ is differentiable in t: $\frac{\partial K(t,s)}{\partial t} = a(t)K(t,s)$ since $K(t,s) = \boldsymbol{E}X_tX_s$, it holds

$$\begin{aligned} K(t,s) = \boldsymbol{E}(X_sX_t) &= \boldsymbol{E}X_s\Big(X_s + \int_s^t a(u)X_u du + \int_s^t b(u)dW'_u\Big) \\ &= \boldsymbol{E}X_s\Big(X_s + \int_s^t a(u)X_u du\Big) \\ &= K(s,s) + \int_s^t a(u)K(s,u)du \end{aligned}$$

and the results.

Now, substituting the right hand side of (6.10) to the Wiener-Hopf equation, we find

$$A(s)K(t,s) \equiv \int_0^t \varphi(t,u)G(u,u)A(s)K(u,s)A(u)du + \varphi(t,s)G(s,s)B^2(s). \tag{6.11}$$

and, differentiating both sides of the above identity in t, we get

$$\begin{aligned} &[a(t) - G(t,t)A(t)]A(s)K(t,s) \equiv \\ &+ \int_0^t \frac{\partial\varphi(t,u)}{\partial t}G(u,u)A(s)K(s,u)A(u)du + \frac{\partial\varphi(t,s)}{\partial t}G(s,s)B^2(s). \end{aligned} \tag{6.12}$$

Applying the right hand side of (6.11) in (6.12) instead of $A(s)K(t,s)$, we obtain

$$\begin{aligned} &[a(t) - G(t,t)A(t)]\Big(\int_0^t \varphi(t,u)G(u,u)A(s)K(s,u)A(u)du + \varphi(t,s)G(s,s)B^2(s)\Big) \\ &\equiv \int_0^t \frac{\partial\varphi(t,u)}{\partial t}G(u,u)A(s)K(s,u)A(u)du + \frac{\partial\varphi(t,s)}{\partial t}G(s,s)B^2(s). \end{aligned} \tag{6.13}$$

By (6.9) $G(t,t)A(t) = \frac{A^2(t)P(t)}{B^2(t)}$ and so we claim that the identity (6.13) may hold provided that for $\varphi_s^s = 1$ and for $u > s$

$$\frac{d\varphi_s^u}{du} = \Big[a(u) - \frac{P(u)A^2(u)}{B^2(u)}\Big]\varphi_s^u. \tag{6.14}$$

Further, by the solution uniqueness of Wiener-Hopf equation we have

$$G(t,s) = \frac{P(t)A(t)}{B^2(t)}\varphi_s^t.$$

It is readily to check that that (home work)

$$\varphi_s^t = \frac{\varphi_0^t}{\varphi_0^s}. \tag{6.15}$$

Then $\widehat{X}_t = \int_0^t G(t,s)dY_s = \int_0^t \varphi_s^t \frac{P(s)A(s)}{B^2(s)} dY_s = \varphi_0^t \int_0^t \frac{1}{\varphi_0^s} \frac{P(s)A(s)}{B^2(s)} dY_s$. From this, by the Itô it holds (home work)

$$d\widehat{X}_t = a(t)\widehat{X}_t dt + \frac{P(t)A(t)}{B^2(t)}\Big[dY_t - A(t)\widehat{X}_t dt\Big]. \tag{6.16}$$

Since $P(t) = \boldsymbol{E}(X_t - \widehat{X}_t)^2$, we derive first the Itô equation for $\Delta(t) = X_t - \widehat{X}_t$. The Itô equations for X_t and $\widehat{X}_t$ imply

$$\begin{aligned} \Delta(t) &= \Delta(0) + \int_0^t a(s)\Delta(s)ds + \int_0^t b(s)dW_s' - \int_0^t \frac{P(s)A^2(s)}{B^2(s)}\Delta(s)ds \\ &\quad - \int_0^t \frac{P(t)A(t)}{B(t)} dW_s''. \end{aligned}$$

Further, by the Itô formula we obtain

$$\begin{aligned} \Delta^2(t) &= \Delta^2(0) + 2\int_0^t a(s)\Delta^2(s)ds + 2\int_0^t \Delta(s)b(s)dW_s' + \int_0^t b^2(s)ds \\ &\quad -2\int_0^t \frac{P(s)A^2(s)}{B^2(s)}\Delta^2(s)ds \\ &\quad -2\int_0^t \frac{P(s)A(s)}{B(s)}\Delta(s)dW_s'' + \int_0^t \frac{P^2(s)A^2(s)}{B^2(s)}ds. \end{aligned}$$

Taking the expectation from both parts of the latter equality, we get (home work)

$$P(t) = P(0) + 2\int_0^t a(s)P(s)ds + \int_0^t b^2(s)ds - \int_0^t \frac{P^2(s)A^2(s)}{B^2(s)}ds. \tag{6.17}$$

Home work

1. Derive (6.4). **Hint:** Taking $\widehat{X}_0 = c_1 + c_2 Y_0$ minimize in c_1, c_2 the mean square error $\boldsymbol{E}(X - \widehat{X}_0)^2$.
2. Show that under $\boldsymbol{E}X_0 = 0$ it holds $\boldsymbol{E}X_t \equiv 0$.
3. To prove (6.15).
4. Derive (6.16) and (6.17).