6. KALMAN-BUCY FILTER


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 $\{Y_s, s \leq t\}$, that is any filtering estimate $\hat{X}_t$ is a functional of $Y^t_0 = \{Y_s, 0 \leq s \leq t\}$. The more popular optimal in the mean square sense filtering estimate is the conditional expectation

$$\hat{X}_t = \mathbb{E}(X_t|Y^t_0)$$

minimizing the mean square error

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

The conditional expectation $\mathbb{E}(X_t|Y^t_0)$, as a function of $Y^t_0$, may have too complicated structure. Therefore, especially in engineering applications, it makes sense to use “a linear function of $Y^t_0$“ minimizing the mean square error $P(t)$ in a class of linear filtering estimates. The optimal linear estimate $\hat{X}_t$ is the orthogonal projection on a linear space generated by $Y^t_0$. 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 $W''_t$.

$$dX_t = a(t)X_t dt + b(t)dW'_t$$
$$dY_t = A(t)X_t dt + B(t)dW''_t$$ (6.1)

\[^1\text{the linear space is generated by any linear combination of } Y_t, \text{ “+ constant” and their closure in the mean square sense} \]
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

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

(6.2)

6.2.1. Description of the Kalman-Bucy filter.

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

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

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

subject to the initial conditions:

$$\hat{X}_0 = E X_0 + \frac{\text{cov}(X_0,Y_0)}{\text{cov}(Y_0,Y_0)} (Y_0 - EY_0)$$

$$P(0) = \text{cov}(X_0,X_0) - \frac{\text{cov}^2(X_0,Y_0)}{\text{cov}(Y_0,Y_0)}.$$

(6.4)

6.2.2. Derivation of filtering equations.

(6.4) is homework.

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 $\hat{X}_t = L(Y_t)$. To describe a structure of $L(Y_t)$, introduce partitions of the interval $[0, t]: 0 = \tau_0^0 < \tau_1^0 < \ldots < \tau_n^0 = t$ so that $\lim_{n \to \infty} \max_{j} |\tau_{j+1}^n - \tau_j^0| = 0$ and find the optimal in the mean square filtering estimate of $X_t$ given $\{Y_{t_j^0}, j = 0, \ldots, n\}$.

Since $\{Y_{t_j^0}, j = 0, \ldots, n\}$ and $\{Y_0, Y_{t_j^0} - Y_{t_{j-1}^0}, j = 1, \ldots, n\}$ contain the same ”information”, we have

$$L^n(Y_0^t) = E(X_t | Y_0, Y_{t_j^0} - Y_{t_{j-1}^0}, j = 1, \ldots n)$$

and, whereas $L^n(Y_t)$ is a linear function of arguments $Y_0, Y_{t_j^0} - Y_{t_{j-1}^0}, j = 1, \ldots n$, we have

$$L^n(Y_0^t) = G_0^n(t) + G_1^n(t)Y_0 + \sum_j G^n(t, t_j^n)[Y_{t_j^n} - Y_{t_{j-1}^0}],$$

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 $E X_0 = 0$ and $\text{cov}(X_0, Y_0) = 0$. Notice that (6.4) provides $G^n(0) \equiv 0$, $G^n_1(0) \equiv 0$. Hence, we have

$$\mathcal{L}^n(Y^t_0) = \sum_j G^n(t, t^n_j)[Y^n_{t^n_j} - Y^n_{t^n_{j-1}}] = \int_0^t G^n(t, s)dY_s,$$

where $G^n(t, s) = G(t, t^n_j)$, $t^n_j \leq s < t^n_{j+1}$.

For further convenience, we choose partitions

$$\{t^n_1, \ldots, t^n_n\} \subseteq \{t^{n+1}_1, \ldots, t^{n+1}_n\}, \, n \geq 1,$$

so that

$$\{Y_0, Y^n_{t^n_j} - Y^n_{t^n_{j-1}}, j = 1, \ldots, n\} \uparrow \{Y_0 = 0, Y_s, s \leq t\}.$$

Then, by the well known property of the conditional expectation

$$\lim_{n \to \infty} \mathbb{E} \left( X_t \mid Y^n_{t^n_j} - Y^n_{t^n_{j-1}}, j = 1, \ldots, n \right) = \mathbb{E} \left( X_t \mid Y_0, Y_s, s \leq t \right),$$

that is

$$\mathbb{E}(X_t|Y^t_0) = \lim_{n \to \infty} \mathcal{L}^n(Y^t_0).$$

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

$$\mathbb{E}(X_t|Y^t_0) = \int_0^t G(t, s)dY_s. \quad (6.5)$$

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

$$\lim_{n \to \infty} \mathcal{L}^n(Y^t_0) \text{ exists} \iff \lim_{n, m \to 0} \mathbb{E} \left( \mathcal{L}^n(Y^t_0) - \mathcal{L}^m(Y^t_0) \right)^2 = 0.$$

On the other hand, since $\mathcal{L}^n(Y^t_0)$ is the linear function of $Y^t_0$, we have

$$\mathcal{L}^n(Y^t_0) - \mathcal{L}^m(Y^t_0) := \int_0^t \left[ G^n(t, s) - G^n(t, s) \right]dY_s$$

and

$$\lim_{n, m \to 0} \mathbb{E} \left( \int_0^t \left[ G^n(t, s) - G^n(t, s) \right]dY_s \right)^2 = 0. \quad (6.6)$$

Therefore, taking into account the independence of $(X_t)$ and $(W''_t)$ and denoting by $K(s', s'') = \mathbb{E}(X_s X_{s''})$ (see Homework 2.), write
\[
\mathbb{E} \left( \int_0^t [G^n(t, s) - G^m(t, s)] dY_s \right)^2 \\
= \mathbb{E} \left( \int_0^t [G^n(t, s) - G^m(t, s)] [A(s)X_s ds + B(s) dW_s'''] \right)^2 \\
= \mathbb{E} \left( \int_0^t [G^n(t, s) - G^m(t, s)] A(s)X_s ds \right)^2 \\
+ \mathbb{E} \left( \int_0^t [G^n(t, s) - G^m(t, s)] B(s) dW_s''' \right)^2 \\
= \int_0^t \int_0^t [G^n(t, s') - G^m(t, s')] \left[ G^n(t, s'') - G^m(t, s'') \right] K(s', s'') ds' ds'' \\
+ \int_0^t [G^n(t, s) - G^m(t, s)]^2 B^2(s) ds.
\]

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 \(B^2(s) \geq c > 0\), 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,

\[
\lim_{n \to \infty} \int_0^t G^n(t, s) A(s) X_s ds = \int_0^t G(t, s) A(s) X_s ds \\
\lim_{n \to \infty} \int_0^t G^n(t, s) B(s) dW_s''' = \int_0^t G(t, s) B(s) dW_s''' .
\]

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 \(s \leq t\)

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

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
\[ \int_0^t f(t, s)dY_s \text{ with deterministic (bounded) function } f(t, s), \text{ it holds} \]
\[ E \left( X_t - \int_0^t G(t, s)dY_s \right) \int_0^t f(t, s)dY_s = 0. \]

The latter equality implies
\[ E \left( X_t \int_0^t f(t, s)dY_s \right) = E \left( \int_0^t G(t, s)dY_s \int_0^t f(t, s)dY_s \right). \]

Write, keeping in mind the independence of \((X_t)\) and \((W'')\),
\[ EX_t \int_0^t f(t, s)dY_s = EX_t \left( \int_0^t f(t, s)A(s)X_sds + \int_0^t f(t, s)B(s)dW''_s \right) \]
\[ = EX_t \int_0^t f(t, s)A(s)X_sds \]
\[ = \int_0^t f(t, s)A(s)E(X_tX_s)ds \]
\[ = \int_0^t f(t, s)A(s)K(t, s)ds \quad (6.7) \]

and
\[ E \int_0^t G(t, s)dY_s \int_0^t f(t, s)dY_s = E \left( \int_0^t G(t, s)A(s)X_sds + \int_0^t G(t, s)B(s)dW''_s \right) \]
\[ \times \left( \int_0^t f(t, s)A(s)X_sds + \int_0^t f(t, s)B(s)dW''_s \right) \]
\[ = E \int_0^t G(t, s)A(s)X_sds \int_0^t f(t, s)A(s)X_sds \]
\[ + E \int_0^t f(t, s)B(s)dW''_s \int_0^t G(t, s)B(s)dW''_s \]
\[ = \int_0^t \int_0^t f(t, s)G(t, u)A(u)K(s, u)A(s)duds \]
\[ + \int_0^t f(t, s)G(t, s)B^2(s)ds. \quad (6.8) \]

(6.7) and (6.8) provide the equality
\[ \int_0^t f(t, s) \left\{ A(s)K(t, s) - \int_0^t G(t, u)A(u)K(s, u)A(s)du \right. \]
\[ \left. - f(t, s)G(t, s)B^2(s) \right\} ds = 0. \]
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

\[
\int_0^t \Delta(t,s)A(s)K(s,u)A(u)du + \Delta(t,u)B^2(u) = 0
\]

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 \( \Delta^2(t,u) = 0, u \leq t \).

**Existence of solution of the Wiener-Hopf equation.**

1. Show first that

\[
G(t,t) = \frac{P(t)A(t)}{B^2(t)}, \quad (6.9)
\]

\[
G(t,t)B^2(t) = K(t,t)A(t) - \int_0^t G(t,u)A(t)K(t,u)A(u)du
\]

\[
= EX^2_tA(t) - \int_0^t G(t,u)A(t)E(X_tX_u)A(u)du
\]

\[
= E\left([X_t - \int_0^t G(t,u)A(u)X_u du]X_tA(t)\right)
\]

\[
= E\left([X_t - \int_0^t G(t,u)A(u)X_u du - \int_0^t G(t,u)B(u)dW_u]X_tA(t)\right)
\]

\[
= E\left([X_t - \int_0^t G(t,u)dY_u]X_tA(t)\right)
\]

\[
= E\left([X_t - \hat{X}_t]X_tA(t) = [X_t - \hat{X}_t][X_t - \hat{X}_t]A(t)
\]

\[
+ [X_t - \hat{X}_t]\hat{X}_tA(t)
\]

\[
= P(t)A(t) + E\left([X_t - \hat{X}_t]\hat{X}_tA(t)\right)
\]

and since \( E\left([X_t - \hat{X}_t]\hat{X}_t\right) = 0 \) we obtain \( G(t,t)B^2(t) = P(t)A(t) \) and (6.9).
2. For \( s < t \), we will find a solution in a form
\[
G(t, s) = \varphi_s G(s, s)
\]
with smooth in \( t \) function \( \varphi_s, \varphi_t = 1 \).

Notice that \( K(t, s) \) is differentiable in \( t \): \( \frac{\partial K(t, s)}{\partial t} = a(t)K(t, s) \) since \( K(t, s) = EX_tX_s \), it holds
\[
K(t, s) = EX_s\left(X_s + \int_s^t a(u)X_u du + \int_s^t b(u)dW_u\right)
\]
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
\[
\left[a(t) - G(t, t)A(t)\right]A(s)K(t, s) \equiv \\
+ \int_0^t \frac{\partial \varphi(t, u)}{\partial t}G(u, u)A(s)K(s, u)A(u)du + \frac{\partial \varphi(t, s)}{\partial t}G(s, s)B^2(s). \tag{6.12}
\]

Applying the right hand side of (6.11) in (6.12) instead of \( A(s)K(t, s) \), we obtain
\[
\left[a(t) - G(t, t)A(t)\right]\left( \int_0^t \varphi(t, u)G(u, u)A(s)K(s, u)A(u)du + \varphi(t, s)G(s, s)B^2(s) \right) \\
\equiv \int_0^t \frac{\partial \varphi(t, u)}{\partial t}G(u, u)A(s)K(s, u)A(u)du + \frac{\partial \varphi(t, s)}{\partial t}G(s, s)B^2(s). \tag{6.13}
\]

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 \( \varphi_s = 1 \) and for \( u > s \)
\[
\frac{d\varphi_s^u}{du} = \left[a(u) - \frac{P(u)A^2(u)}{B^2(u)}\right]\varphi_s^u. \tag{6.14}
\]

Further, by the solution uniqueness of Wiener-Hopf equation we have
\[
G(t, s) = \frac{P(t)A(t)}{B^2(t)}\varphi_s^t.
\]
The Kalman filter.

It is readily to check that that (home work)
\[ \phi_t = \phi_0. \] (6.15)

Then \( \hat{X}_t = \int_0^t G(t, s) dY_s = \int_0^t \phi_s \frac{P(s)A(s)}{B^2(s)} dY_s = \phi_0 \int_0^t \frac{1}{\phi_0} \frac{P(s)A(s)}{B^2(s)} dY_s. \) From this, by the Itô it holds (home work)
\[
d\hat{X}_t = a(t)\hat{X}_t dt +\frac{P(t)A(t)}{B^2(t)} \left[ dY_t - A(t)\hat{X}_t dt \right]. \] (6.16)

Since \( P(t) = \mathbf{E}(X_t - \hat{X}_t)^2, \) we derive first the Itô equation for \( \Delta(t) = X_t - \hat{X}_t. \) The Itô equations for \( X_t \) and \( \hat{X}_t \) imply
\[
\Delta(t) = \Delta(0) + \int_0^t a(s)\Delta(s) ds + \int_0^t b(s)dW'_s - \int_0^t \frac{P(s)A^2(s)}{B^2(s)} \Delta(s) ds \\
- \int_0^t \frac{P(t)A(t)}{B(t)} dW''_s.
\]

Further, by the Itô formula we obtain
\[
\Delta^2(t) = \Delta^2(0) + 2 \int_0^t a(s)\Delta^2(s) ds + 2 \int_0^t \Delta(s)b(s)dW'_s + \int_0^t b^2(s)ds \\
- 2 \int_0^t \frac{P(s)A^2(s)}{B^2(s)} \Delta(s) ds \\
- 2 \int_0^t \frac{P(t)A(t)}{B(t)} \Delta(s) dW''_s + \int_0^t \frac{P^2(s)A^2(s)}{B^2(s)} ds.
\]
Taking the expectation from both parts of the latter equality, we get (home work)
\[
P(t) = P(0) + 2 \int_0^t a(s)P(s)ds + \int_0^t b^2(s)ds - \int_0^t \frac{P^2(s)A^2(s)}{B^2(s)} ds. \] (6.17)

**Home work**

1. Derive (6.4). **Hint:** Taking \( \hat{X}_0 = c_1 + c_2Y_0 \) minimize in \( c_1, c_2 \) the mean square error \( \mathbf{E}(X - \hat{X}_0)^2. \)
2. Show that under \( \mathbf{E}X_0 = 0 \) it holds \( \mathbf{E}X_t \equiv 0. \)
3. To prove (6.15).
4. Derive (6.16) and (6.17).