# 8. TRANSMISSION OF A GAUSSIAN SIGNAL THROUGH A CHANNEL WITH NOISELESS FEEDBACK. OPTIMAL CODING AND DECODING

8.1. **Transmission of Gaussian random variable.**

Let $X$ be a Gaussian random variable with

$$\boldsymbol{E}X = m \quad \text{and} \quad \boldsymbol{E}(X-m)^2 = P > 0.$$

We consider $X$ as a signal which has to be transmitted through a noise channel. Assume also that the exit $(Y_t)_{t\geq 0}$ of the noisy channel is observable.

Three parameters: the time $t$, the signal $X$, and the observation $Y_0^t$ might me used for creating of a code $A(t,X,Y_0^t)$ for $X$. In other words, at the time $t$, instead of $X$, we transmit a new signal $A(t,X,Y_0^t)$ generated by $t$, $X$, and observation $Y_0^t = \{Y_s, 0 \leq s \leq t\}$. The transmitted signal is interrupted by Gaussian white noise: $A(t,X,Y_0^t)$ + "white noise", i.e.

$$dY_t = A(t,X,Y_0^t)dt + \sigma dW_t, \; Y_0 = 0,$$

where $W_t$ is Wiener process independent of $X$ and $\sigma$ is the noise intensity.

Our task now is to choose a coding functional $A(t,X,Y_0^t)$ such that the signal $X$ is transmitted with a good accuracy. It would be noted that without any restrictions such a setting of problem is useless. In fact, with $N$, being essentially larger than $\sigma^2$, taking $A(t,X,Y_0^t) = NX$ and letting the filtering estimate

$$\widehat{X}_t = \frac{Y_t}{tN},$$

we find that

$$\widehat{X}_t = X + \frac{bW_t}{tN}$$

and therefore the mean square error is small

$$\boldsymbol{E}(X-\widehat{X}_t)^2 = \boldsymbol{E}\Big(\frac{\sigma W_t}{tN}\Big)^2 = \frac{\sigma^2}{tN^2} \text{ (is small).}$$

The choice of a large parameter $N$ is unrealistic since an energy which has to be transmitted through the channel is too large: $\boldsymbol{E}(NX)^2 = N^2(m^2+P)$ what is very restrictive in practice.

To avoid such type of unrealistic solution, we require that for some $p > 0$ and any $t > 0$

$$\frac{1}{t}\int_0^t \boldsymbol{E}A^2(s,X,Y_0^s)ds \leq p.$$

If a coding functional $A$ is chosen, then the optimal in the mean square sense decoding procedure consists in computing of the optimal estimate $\widehat{X}_t = \boldsymbol{E}(X|Y_0^t)$.

Due to a non linear character of the equation for $Y_t$, the pair $X, Y_t$ forms non Gaussian random process, so that an explicit computation of $\widehat{X}_t = \boldsymbol{E}(X|Y_0^t)$ is problematic. Therefore, we restrict ourselves by a class of coding functionals linear in $X$, that is we consider [1] $A(t, X, Y_0^t) = [v(t, Y_0^t) + V(t, Y_0^t)X]$, i.e.

$$dY_t = [v(t, Y_0^t) + V(t, Y_0^t)X]dt + dW_t \tag{8.1}$$

Despite it that case $X, Y_t$ also is non Gaussian object, the conditional distribution $P(X \leq x|Y_0^t)$ is Gaussian (a.s.) for every $t > 0$.

Indeed, we deal with the model

$$\begin{aligned} dX_t &= 0 \\ dY_t &= [v(t, Y_0^t) + V(t, Y_0^t)X]dt + \sigma dW_t, \end{aligned} \tag{8.2}$$

where $X_0 = X, Y_0 = 0$.

From Lecture 7 (**Conditionally Gaussian filter**) it holds that $\widehat{X}_t = \boldsymbol{E}(X|Y_0^t)$ and $P(t) = \boldsymbol{E}\big((X - \widehat{X}_t)^2|Y_0^t\big)$ are defined by the conditionally Gaussian filter (CGF):

$$\begin{aligned} d\widehat{X}_t &= \frac{P(t)V(t, Y_0^t)}{\sigma^2}[dY_t - (v(t, Y_0^t) + V(t, Y_0^t)\widehat{X}_t)dt] \\ \dot{P}(t) &= -\frac{P^2(t)V(t, Y_0^t)}{\sigma^2}, \end{aligned} \tag{8.3}$$

subject to $\widehat{X}_0 = \boldsymbol{E}X$ and $P(0) = P$. In additional

$$\overline{W}_t = \int_0^t \frac{dY_s - \big(v(s, Y_0^s) + V(s, Y_0^s)\widehat{X}_t\big)}{\sigma^2} ds \tag{8.4}$$

is the innovation Wiener process.

8.2. **The choice of optimal** $v(t, Y_0^t), V(t, Y_0^t)$**.**

We have to choose $v(t, Y_0^t), V(t, Y_0^t)$ under

$$t \geq \frac{1}{p}\int_0^t \boldsymbol{E}[v(s, Y_0^s) + V(s, Y_0^s)X_s]^2 dt, \ \ t > 0. \tag{8.5}$$

Show now that such the choice guarantees the minimal value of $P(t), t \geq 0$. In fact, the differential equation for $P(t)$ can be rewritten into

$$\frac{\dot{P}(t)}{P(t)} = -\frac{P(t)V^2(t, Y_0^t)}{\sigma^2}$$

what implies

$$P(t) = P\exp\Big(-\frac{1}{\sigma^2}\int_0^t P(s)V^2(s, Y_0^s)ds\Big). \tag{8.6}$$

Therefore $\int_0^t P(s)V^2(s, Y_0^s)ds$ has to be minimized by a relevant choice of $V$.

[1] $v(t, Y_0^t)$ are $V(t, Y_0^t)X]$ have to chosen such that the Itô equation (8.1) obeys the unique solution

To choose the required $V$, write

$$
\begin{aligned}
pt &\geq \int_0^t \boldsymbol{E}[v(s,Y_0^s)+V(s,Y_0^s)X_s]^2 dt \\
&= \int_0^t \boldsymbol{E}[v(s,Y_0^t)+V(s,Y_0^s)\widehat{X}_s+V(s,Y_0^s)(X_s-\widehat{X}_s)]^2 ds \\
&= \int_0^t \boldsymbol{E}[(v(s,Y_0^s)+V(s,Y_0^s)\widehat{X}_s)^2+V^2(s,Y_0^s)P(s)]ds \\
&\geq \int_0^t \boldsymbol{E}V^2(s,Y_0^s)P(s)ds,
\end{aligned}
$$

and thus

$$
\int_0^t \boldsymbol{E}V^2(s,Y_0^s)P(s)ds \leq tp, \ \forall t>0. \tag{8.7}
$$

Let us find now a lower bound for $\boldsymbol{E}P(t)$. With the help of (8.6) and (8.7) we find

$$
\begin{aligned}
\boldsymbol{E}P(t) &= P\boldsymbol{E}\exp\Big(-\frac{1}{\sigma^2}\int_0^t P(s)V^2(s,Y_0^s)ds\Big) \\
&\geq P\exp\Big(-\frac{1}{\sigma^2}\int_0^t \boldsymbol{E}[P(s)V^2(s,Y_0^s)]ds\Big) \ \text{(Jensen inequality)} \\
&\geq P\exp\Big(-\frac{pt}{\sigma^2}\Big).
\end{aligned} \tag{8.8}
$$

We show now that the lower bound given in (8.8) is attainable on

$$
v^*(s,Y_0^t)=-V^*(s,Y_0^s)\widehat{X}_s \quad \text{and} \quad V^*(t,Y_0^t)=\sqrt{\frac{p}{P^*(t)}},
$$

where

$$
\dot{P}^*(t)=-\frac{(V^*(t,Y_0^t)P^*(t))^2}{\sigma^2}\equiv -\frac{pP^*(t)}{\sigma^2}
$$

so that $\boldsymbol{E}P^*(t)\equiv P^*(t)\equiv Pe^{-\frac{pt}{\sigma^2}}$ and

$$
V^*(t,Y_0^t)=\sqrt{\frac{p}{P}}\exp\Big(\frac{pt}{2\sigma^2}\Big).
$$

Thus, the optimal coding-decoding scheme is given by

$$
d\widehat{X}_t^*=\frac{P^*(t)V^*(t,Y_0^t)}{\sigma^2}dY_t^*=\sqrt{\frac{pP}{\sigma^2}}e^{-pt/2\sigma^2}dY_t^*,
$$

where

$$
dY_t^*=\sqrt{\frac{p}{P}}e^{\frac{pt}{2\sigma^2}}\Big(X-\widehat{X}_t^*\Big)dt+\sigma dW_t.
$$

8.3. **Transmission of Gaussian random signal.**

Consider the same problem with a transmitted signal $(X_t)_{t\ge 0}$ being a Gaussian random process process defined by the Itô equation (with respect to a Wiener process $(W'_t)_{t\ge 0}$, $a(t)$ and $b(t)$ are continuous functions):

$$dX_t = a(t)X_t dt + b(t)dW'_t$$

subject to Gaussian initial condition $X_0$ with $\boldsymbol{E}X_0 = m$ and $\boldsymbol{E}(X_0 - m)^2 = P$. As previously, a coder is taken in the form $A(t, X_t, Y_0^t) = v(t, Y_0^t) + V(t, Y_0^t)X_t$, i.e.

$$dY_t = [v(t, Y_0^t) + V(t, Y_0^t)X]dt + \sigma dW_t,\ Y_0 = 0,$$

where $W_t$ is a Wiener process independent of $W'_t$. Functionals $v(t, Y_0^t)$ and $A(t, Y_0^t)$ are subjects to the constrain: for some $p > 0$ and any $t > 0$

$$\boldsymbol{E}(v(t, Y_0^t) + V(t, Y_0^t)X_t)^2 \le p. \tag{8.9}$$

The conditionally Gaussian filter defines $\widehat{X}_t = \boldsymbol{E}(X_t|Y_0^t)$ and $P(t) = \boldsymbol{E}\big((X_t - \widehat{X}_t)^2|Y_0^t\big)$:

$$\begin{aligned} d\widehat{X}_t &= a(t)\widehat{X}_t dt + \frac{P(t)V(t, Y_0^t)}{\sigma^2}\Big(dY_t - [v(t, Y_0^t) + V(t, Y_0^t)\widehat{X}_t]dt\Big) \\ \dot{P}(t) &= 2a(t)P(t) + \sigma^2(t) - \frac{P^2(t)A^2(t, Y_0^t)}{\sigma^2} \end{aligned} \tag{8.10}$$

subject to the initial conditions $\widehat{X}_0 = m$ and $P(0) = P$.

To find the optimal $v^*$ and $V^*$, we apply the same method which has been used for finding the optimal coding functionals for the case of transmitting Gaussian random variable $X$. To this end, let us transform the second equation in (8.10) into the form

$$\dot{P}(t) = \Big(2a(t) - \frac{P(t)V^2(t, Y_0^t)}{\sigma^2}\Big)P(t) + \sigma^2(t)$$

which allows to arrive at

$$\begin{aligned} P(t) &= P(0)\exp\Big\{\int_0^t \Big(2a(s) - \frac{P(s)V^2(s, Y_0^s)}{\sigma^2}\Big)ds\Big\} \\ &\quad + \int_0^t b^2(s)\exp\Big\{\int_s^t (2a(s') - \frac{P(s')V^2(s', Y_0^{s'}}{\sigma^2}))ds'\Big\}ds. \end{aligned} \tag{8.11}$$

We use now the restriction given in (8.9):

$$\begin{aligned}
p &\geq \boldsymbol{E}(v(t,Y_0^t)+V(t,Y_0^t)X_t)^2\\
&= \boldsymbol{E}(v(t,Y_0^t)+V(t,Y_0^t)\widehat{X}_t+V(t,Y_0^t)[X_t-\widehat{X}_t])^2\\
&= \boldsymbol{E}(v(t,Y_0^t)+V(t,Y_0^t)\widehat{X}_t)^2+P(t)V^2(t,Y_0^t)\\
&\geq \boldsymbol{E}P(t)V^2(t,Y_0^t).
\end{aligned}$$

The inequality obtained jointly with (8.11) imply

$$\begin{aligned}
\boldsymbol{E}P(t) &= P(0)\boldsymbol{E}\exp\Big\{\int_0^t\Big(2a(s)-\frac{P(s)V^2(s,Y_0^s)}{\sigma^2}\Big)ds\Big\}\\
&\quad+\int_0^t b^2(s)\boldsymbol{E}\exp\Big\{\int_s^t\Big(2a(s')-\frac{P(s')V^2(s',Y_0^{s'})}{\sigma^2}\Big)ds'\Big\}ds\\
&\geq P(0)\exp\Big\{\int_0^t\Big(2a(s)-\boldsymbol{E}\Big[\frac{P(s)V^2(s,Y_0^s)}{\sigma^2}\Big]\Big)ds\Big\}\ \text{(Jensen ineq.)}\\
&\quad+\int_0^t b^2(s)\exp\Big\{\int_s^t\Big(2a(s')-\boldsymbol{E}\Big[\frac{P(s')V^2(s',Y_0^{s'})}{\sigma^2}\Big]\Big)ds'\Big\}ds\ \text{(Jensen ineq.)}\\
&\geq P(0)\exp\Big\{\int_0^t(2a(s)-p)ds\Big\}\\
&\quad+\int_0^t b^2(s)\exp\Big\{\int_s^t\Big(2a(s')-\frac{p}{\sigma^2}\Big)ds'\Big\}ds. \qquad (8.12)
\end{aligned}$$

Evidently this lower bound is attainable on

$$v^*(t,Y_0^t)=-\widehat{X}_tV^*(t,Y_0^t)\quad\text{and}\quad V^*(t,Y_0^t)=\sqrt{\frac{p}{P^*(t)}}.$$

That choice of $V^*(t,Y_0^t)$ transforms nonlinear differential Ricatti equation into a linear one

$$\dot{P}^*(t)=\Big[2a(t)-\frac{p}{\sigma^2}\Big]P^*(t)+b^2(t)$$

subject to $P(0)=P$.

Thus the optimal decoding scheme is given by

$$\begin{aligned}
d\widehat{X}_t^* &= a(t)\widehat{X}_t^*dt+\frac{\sqrt{pP^*(t)}}{\sigma^2}dY_t^*\\
dY_t^* &= \sqrt{\frac{p}{P^*(t)}}[X_t-\widehat{X}_t^*]dt+\sigma dW_t.
\end{aligned}$$

**Remark.** It is shown in R.S. Liptser and A.N. Shiryaev "Statistics of Random Processes, II. Application" (Ch. 16, §16.3) that in both examples the coding functional

$v^*(t,Y_0^t)+V^*(t,Y_0^t)X_t$ is optimal among functionals of non linear in $X$: $A(t,X,Y_0^t)$ if only the above-mentioned energy type restrictions are valid.

8.4. **Example.**

Assume $a(t)=-1$ that is $dX_t=-X_tdt+dW_t'$. Apply the coding $V^\circ(t)X_t$, where

$$V^\circ(t)=\sqrt{\frac{p}{\boldsymbol{E}X_t^2}}.$$

Assume that $\boldsymbol{E}X_0=0$. Then, using the Itô formula for $X_t^2$, we find

$$dX_t^2=[-2X_t^2+1]dt-2X_tdW_t'$$

and, after taking the expectation, we arrive to the differential equation for $Q(t)=\boldsymbol{E}X_t^2$:

$$\dot{Q}_t=-2Q_t+1.$$

Therefore $Q_t\to 1/2$ for $t\to\infty$ and so for large $t$ we have $V^\circ(t)\approx\sqrt{2p}$. At the same time the decoding error $P(t)$, corresponding to $V^\circ(t)$, is defined by the Ricatti equation

$$\dot{P}(t)=-2P(t)+1-(V^\circ(t))^2P^2(t)$$

and so for large values of $t$, $P(t)\approx\frac{\left[\sqrt{1+4p}-1\right]}{2p}$.

Therefore

$$\frac{P^*(\infty)}{P(\infty)}\sim\begin{cases}\sqrt{\frac{2}{p}}, & p\to\infty\\ 1/2, & p\to 0\end{cases}$$

In other words, feedback coding '$\sqrt{\frac{p}{P^*(t)}}(X_t-\widehat{X}_t^*)$' yields much smaller decoding mean square error a specially for large $p$.

**Home work.**

A random signal $X_n, n=0,1...$, has to be transmitted through a noise channel, that is an output signal $Y_n, n=1,2...$, contains incomplete information on the transmitted signal:

$$Y_{n+1}=A_0(n,Y_0^n)+A_1(n,Y_0^n)X_n+\xi_{n+1},$$

where $\xi_n, n=1,2...$, is i.i.d. Gaussian sequence with $\boldsymbol{E}\xi_1=0$, $\boldsymbol{E}\xi_1^2=1$, $Y_0=0$, and $Y_0^n=(Y_0,Y_1,...,Y_n)$. $A_0(n,Y_0^n)$ and $A_1(n,Y_0^n)$ are coding functionals. Denote by $\widehat{X}_n = \boldsymbol{E}(X_n|Y_0^n)$. A quality of the coding-decoding system is characterized by the mean square value $\boldsymbol{E}(X_n - \widehat{X}_n)^2$.

Under the constrain

$$\boldsymbol{E}\Big(A_0(n,Y_0^n) + A_1(n,Y_0^n)X_n\Big)^2 \le p,\ \ n=0,1,...,$$

where $p$ is some fixed constant, find optimal $A_0(n,Y_0^n)$ and $A_1(n,Y_0^n)$, if

$$X_{n+1} = aX_n + b\eta_{n+1},$$

where $\eta_n, n=1,2,...$ is i.i.d. Gaussian sequence (with $\boldsymbol{E}\eta_1 = 0$, $\boldsymbol{E}\eta_1^2 = 1$) independent of $Y_n, \xi_n, n=1,2,...$

**Hint:** Use the conditionally Gaussian filter

$$\begin{aligned}
\widehat{X}_{n+1} &= a\widehat{X}_n + \frac{aP_nA_1(n,Y_0^n)}{\sigma^2 + P_nA_1^2(n,Y_0^n)}\Big(Y_{n+1} - A_0(n,Y_0^n) - A_1(n,Y_0^n)\widehat{X}\Big)_n, \\
P_{n+1} &= a^2P_n + 1 - \frac{(aP_nA_1(n,Y_0^n))^2}{\sigma^2 + P_nA_1^2(n,Y_0^n)},
\end{aligned}$$

where $P_n = \boldsymbol{E}\Big((X_n - \widehat{X}_n)^2|Y_0^n\Big)$.