9. STOCHASTIC FEEDBACK CONTROL. BELLMAN EQUATION

9.1. **Scalar model.** We consider now a controlled process X_t is defined by a linear Itô equation with respect to Wiener process W_t :

$$dX_t = [aX_t + u(t)]dt + bdW_t \tag{9.1}$$

subject to the fixed initial point $X_0 = x$. Here $u(t)$ is the control affecting the drift ($b\frac{dW_t}{dt}$ is Gaussian white noise with intensity b^2). The process X_t is assumed to be observable, so that the control $u(t)$ at every time value t might be chosen as a function of t, X_0^t, i.e. control

$$u(t) = u(t; X_s, s \le t)(= u(t, X_{[0,t]})), \tag{9.2}$$

is of a feedback type.

As for the deterministic setting ($b = 0$), we consider here the quadratic cost functional (recall $X_0 = x$)

$$J(x, u) = \boldsymbol{E}\Big\{X_T^2 + \int_0^T \big(X_t^2 + (u(t))^2\big)dt\Big\}. \tag{9.3}$$

The abbreviation LQG is used for this controlled model:
L - linear ((9.1) is linear equation);
Q - quadratic cost functional;
G - Gaussian noise.

As for deterministic model, we apply here *Dynamic Programming Method.* The main difference from the deterministic case consists in the using of the Itô stochastic calculus.

Introduce the family of cost functionals parametrized by t, x:

$$V(t, x) = \inf_{u(s, X_{[t,s]}), t \le s \le T} \boldsymbol{E}\Big\{X_T^2 + \int_t^T \big(X_s^2 + (u(s, X_{[t,s]}))^2\big)ds\Big\},$$

where $X_s, t < s \le T$ is the solution of (9.1)

$$dX_s = aX_s ds + u(s, X_{[t,s]})ds + bdW_s$$

subject at the time value t by the initial condition $X_t = x$, and note that $V(0, x)$ is the value of the optimal cost for the original problem. As in the deterministic case, we derive for $V(t, x)$ the Bellman equation, being the partial differential equation of the second order (recall that for the deterministic case the Bellman equation is partial differential equation of the first order) which is main tool for creating the optimal control.

9.1.1. **Principle of optimality. Heuristic method.**

As in the deterministic case, we use the principle of optimality which says: if $\tilde{u}(s, X_{[t+\delta,s]})$, $t+\delta \le s \le T$ is the optimal control for the time interval $[t+\delta, T]$ corresponding to the initial point $X_{t+\delta}$, then the optimal control for the interval $[t, T]$ minimizes the functional

$$\boldsymbol{E}\Big\{\int_t^{t+\delta} \big(X_s^2 + (u(s, X_{[t,s]}))^2\big) ds + V(t+\delta, X_{t+\delta})\Big\}.$$

In other words, for $\delta > 0$ and $t < T - \delta$, we have

$$\begin{aligned}
V(t,x) &= \inf_{u(s,X_{[t,s]}), t\le s\le T} \boldsymbol{E}\Big\{X_T^2 + \int_{t+\delta}^T \big(X_s^2 + (u(s, X_{[t+\delta,s]}))^2\big) ds \\
&\quad + \int_t^{t+\delta} \big(X_s^2 + (u(s, X_{[t,s]}))^2\big) ds\Big\} \\
&= \inf_{u(s,X_{[t,s]}), t\le s\le t+\delta} \boldsymbol{E}\Big\{\int_t^{t+\delta} \big(X_s^2 + (u(s, X_{[t,s]}))^2\big) ds \\
&\quad + \inf_{u(s,X_{[t+\delta,s]}), t+\delta\le s\le T} \Big\{X_T^2 + \int_{t+\delta}^T \big(X_s^2 + (u(s, X_{[t+\delta,s]}))^2\big) ds\Big\} \\
&= \inf_{u(s,X_{[t,s]}), t\le s\le t+\delta} \boldsymbol{E}\Big\{\int_t^{t+\delta} \big(X_s^2 + (u(s, X_{[t,s]}))^2\big) ds + V(t+\delta, X_{t+\delta})\Big\}
\end{aligned}$$

which implies

$$\begin{aligned}
\frac{V(t,x) - V(t+\delta,x)}{\delta} &= \inf_{u(s,X_{[t,s]}), t\le s\le t+\delta} \boldsymbol{E}\Big\{\frac{\int_t^{t+\delta} \big(X_s^2 + (u(s, X_{[t,s]}))^2\big) ds}{\delta} \\
&\quad + \frac{V(t+\delta, X_{t+\delta}) - V(t+\delta, x)}{\delta}\Big\}
\end{aligned}$$

and, in turn, we find

$$\begin{aligned}
\boldsymbol{E}\Big\{V(t+\delta, X_{t+\delta}) - V(t+\delta, x)\Big\} &\approx \frac{\partial V(t,x)}{\partial t}\delta \\
&\quad + \frac{\partial V(t,x)}{\partial x}(ax + u(t,x))\delta \\
&\quad + \frac{1}{2}\frac{\partial^2 V(t,x)}{\partial x^2} b^2 E(W_{t+\delta} - W_t)^2
\end{aligned}$$

and with $\delta \to 0$ we arrive at the partial differential equation

$$-\frac{\partial V(t,x)}{\partial t} = \inf_u \Big\{x^2 + u^2 + \frac{\partial V(t,x)}{\partial x}\big[ax + u\big] + \frac{b^2}{2}\frac{\partial^2 V(t,x)}{\partial x^2}\Big\} \tag{9.4}$$

subject to the boundary condition condition $V(T,x) = x^2$. Equation (9.4) and the function $V(t,x)$ are named the Bellman equation and function respectively.

Write

$$u^*(t,x) = \operatorname*{argmin}_u \Big\{ x^2 + u^2 + \frac{\partial V(t,x)}{\partial x}\big[ax + u\big]\Big\} \tag{9.5}$$

and create the control of the feedback type $u^*(t,X^*_t)$, where X^*_t is the solution of differential equation (9.1) corresponding to the control $u^*(t,X^*_t)$, that is

$$dX^*_t = aX^*_t dt + u^*(t,X^*_t)dt + bdW_t. \tag{9.6}$$

9.1.2. $\boldsymbol{u^*(t,X^*_t)}$ **is the optimal control.**

To show that $u^*(t,X^*_t)$ is the optimal control, let consider the random process $V(t,X^*_t)$. Since the function $V(t,x)$ is continuously differentiable one in t and twice in x, the Itô formula is applicable to $V(t,X^*_t)$:

$$\begin{aligned} dV(t,X^*_t) &= V'_t(t,X^*_t) + V'_x(t,X^*_t)\big[aX^*_t + u^*(t,X^*_t)\big]dt \\ &\quad + V'_x(t,X^*_t)bdW_t \\ &\quad + \frac{1}{2}V''_{xx}(t,X^*_t)b^2 dt. \end{aligned} \tag{9.7}$$

The use of (9.4) provides identity

$$-V'_t(t,x) \equiv x^2 + (u^*(t,x))^2 + V'_x(t,x)\big[ax + u^*(t,x)\big] + \frac{b^2}{2}V''_{xx}(t,x)$$

and so, for $x = X^*_t$ we have

$$\begin{aligned} -V'_t(t,X^*_t) &= (X^*_t)^2 + (u^*(t,X^*_t))^2 + V'_x(t,X^*_t)\big[aX^*_t + u^*(t,X^*_t)\big] \\ &\quad + \frac{b^2}{2}V''_{xx}(t,X^*_t) \end{aligned} \tag{9.8}$$

Combining now (9.7) and (9.8) we find

$$dV(t,X^*_t) = -\big[(X^*_t)^2 + (u^*(t,X^*_t))^2\big]dt + V'_x(t,X^*_t)bdW_t$$

or, in the integral form, taking into account $V(T,x) = x^2$,

$$(X^*_t)^2 = V(T,X^*_T) = V(0,x) - \int_0^T \big[(X^*_t)^2 + (u^*(t,X^*_t))^2\big]dt + \int_0^T V'_x(t,X^*_t)bdW_t.$$

Now, taking the expectation from both parts of the latter equality and taking into account $\boldsymbol{E}\int_0^T V'_x(t,X^*_t)bdW_t = 0$ and $V(0,x)$ is non random, we find

$$V(0,x) = \boldsymbol{E}\Big\{(X_t^*)^2 + \int_0^T \big[(X_t^*)^2 + (u^*(t,X_t^*))^2\big]dt\Big\}. \tag{9.9}$$

Now, let $u(t,X_{[0,t]})$ be any admissible control, where X_t is the corresponding controlled process. By the Itô formula we find

$$\begin{aligned} dV(t,X_t) &= V_t'(t,X_t) + V_x'(t,X_t)\big[aX_t + u(t,X_{[0,t]})\big]dt \\ &\quad + V_x'(t,X_t)bdW_t \\ &\quad + \frac{1}{2}V_{xx}''(t,X_t)b^2dt. \end{aligned} \tag{9.10}$$

On the other hand, for any x (9.4) implies

$$\begin{aligned} -V_t'(t,x) &\le x^2 + (u(t,X_{[0,t]})^2 + V_x'(t,x)\big[ax + u(t,X_{[0,t]})\big] \\ &\quad + \frac{b^2}{2}V_{xx}''(t,x) \end{aligned}$$

and so, for $x = X_t$ we get

$$\begin{aligned} -V_t'(t,X_t) &\le (X_t)^2 + u^2(t,X_{[0,t]}) + V_x'(t,x)\big[aX_t + u(t,X_{[0,t]})\big] \\ &\quad + \frac{b^2}{2}V_{xx}''(t,X_t) \end{aligned} \tag{9.11}$$

Combining now (9.10) and (9.11) we arrive at

$$dV(t,X_t) \ge -\big[(X_t)^2 + u^2(t,X_{[0,t]})\big]dt + V_x'(t,X_t)bdW_t$$

or, in the integral form, taking into account $V(T,x) = x^2$,

$$(X_t)^2 = V(T,X_T) \ge V(0,x) - \int_0^T \big[(X_t)^2 + u^2(t,X_{[0,t]}))\big]dt + \int_0^T V_x'(t,X_t)bdW_t.$$

Taking the expectation from both parts of the last equality and taking into account $\boldsymbol{E}\int_0^T V_x'(t,X_t)bdW_t = 0$ and $V(0,x)$ is non random, we obtain

$$V(0,x) \le \boldsymbol{E}\Big\{X_t^2 + \int_0^T \big[X_t^2 + u^2(t,X_{[0,t]})\big]dt\Big\} = J(x,u). \tag{9.12}$$

(9.12) gives the lower bound for any cost $J(x,u)$. Consequently

$$u^*(t,X_t^*) \quad \text{is the optimal control.} \tag{9.13}$$

Also note that that the optimal control is the Markov control, i.e. the control depending only on the time value t and the value of the controlled process at the time t.

9.1.3. **Creating of the optimal control.**
Due to (9.5)

$$u^*(t,x) = -\frac{1}{2}\frac{\partial V(t,x)}{\partial x}. \tag{9.14}$$

Substituting (9.14) in (9.8) we arrive at the non linear partial differential equation

$$-\frac{\partial V(t,x)}{\partial t} = \Big\{x^2 + \frac{\partial V(t,x)}{\partial x}ax - \frac{1}{4}\Big(\frac{\partial V(t,x)}{\partial x}\Big)^2 + \frac{b^2}{2}\frac{\partial^2 V(t,x)}{\partial x^2}\Big\}. \tag{9.15}$$

As for the deterministic case, we will find the solution of (9.15) in a special form of a quadratic function

$$V(t,x) = \Gamma(t)x^2 + B(t)x + Q(t),$$

where $\Gamma(t)$, $B(t)$, and $Q(t)$ are assumed to be continuously differentiable functions. Since $V(T,x) = x^2$ it is clear that $\Gamma(T) = 1$ and $B(T) = Q(T) = 0$. Also note that

$$\begin{aligned} \frac{\partial V(t,x)}{\partial x} &= 2\Gamma(t)x + B(t) \\ \frac{\partial^2 V(t,x)}{\partial x^2} &= 2\Gamma(t). \end{aligned}$$

Substituting $V(t,x) = \Gamma(t)x^2 + B(t)x + Q(t)$ in (9.15) we find

$$-\Big(\dot\Gamma(t)x^2 + \dot B(t)x + \dot Q(t)\Big) = x^2 + ax\big(2\Gamma(t)x + B(t)\big) - \frac{1}{4}\big(2\Gamma(t)x + B(t)\big)^2 + b^2\Gamma(t)$$

or, what is equivalent due to an arbitrariness of x,

$$\begin{aligned} -\dot\Gamma(t) &= 1 + 2a\Gamma(t) - \Gamma^2(t) \\ -\dot B(t) &= aB(t) - 2\Gamma(t)B(t) \\ -\dot Q(t) &= -\frac{1}{4}B^2 - b^2\Gamma(t). \end{aligned}$$

The above-mentioned boundary conditions imply $B(t) \equiv 0$ and $Q(t) = b^2\int_t^T \Gamma(s)ds$. Hence

$$V(t,x) = \Gamma(t)x^2 + b^2\int_0^T \Gamma(s)ds,$$

where $\Gamma(t)$ is a solution of Riccati's equation

$$-\dot\Gamma(t) = 1 + 2a\Gamma(t) - \Gamma^2(t), \ 0 \le t < T \tag{9.16}$$

subject to the boundary condition $\Gamma(T)=1$.

Now, we have

$$u^*(t,x) = -\Gamma(t)x.$$

Thus, the optimal controlled model is described by the differential equation:

$$dX^*_t = \big(a-\Gamma(t)\big)X^*_t dt + bdW_t.$$

9.2. **Remark.**

The function $(u^*(t,x))$ is independent of b (the intensity of white noise). So for the stochastic case ($b \neq 0$) this function is the same as for the deterministic case ($b=0$). The value of the optimal cost functional is larger for the stochastic case.

9.3. **The Bellman equation for non linear model.**

Now, we consider a non linear stochastic controlled model

$$dX_t = [a(t,X_t) + A(t,X_t)u(t,X_{[0,t]})]dt + bdW_t$$

subject to the initial point $X_0 = x$. The Bellman equation, corresponding to the cost functional

$$J(x) = \boldsymbol{E}\Big\{F(X_t) + \int_0^T \big(h(X_t) + u^2(t,X_{[0,t]})\big)dt\Big\},$$

is the following:

$$-\frac{\partial V(t,x)}{\partial t} = \inf_u \Big\{h(x) + u^2 + \frac{\partial V(t,x)}{\partial x}\big[a(t,x) + A(t,x)u\big] + \frac{b^2}{2}\frac{\partial^2 V(t,x)}{\partial x^2}\Big\}.$$

subject to the boundary condition $X(T,x) = F(x)$. Therefore,

$$\begin{aligned} u^*(t,x) &= \operatorname{argmin}_u \Big\{u^2 + \frac{\partial V(t,x)}{\partial x}A(t,x)u\Big\} \\ &= -\frac{1}{2}V'_x(t,x)A(t,x) \end{aligned}$$

and $u^*(t,X^*_t)$ can be used as the optimal control provided that the Itô equation

$$dX^*_t = [a(t,X^*_t) + A(t,X^*_t)u^*(t,X^*_t)]dt + bdW_t$$

obeys solution.

Home work.

Let $X_k, k=0,1,...,N$ be a controlled sequence defined by linear recursion

$$X_{k+1} = aX_k + cu_k + b\xi_k,$$

where $u_k, k = 0, 1..., N-1$ is the control, $\xi_k, k = 1, 2...$ is i.i.d. sequence of random variables with $\boldsymbol{E}\xi_1 = 0$, $\boldsymbol{E}\xi_1^2 = 1$. For each k, a control u_k is chosen such that to minimize the cost (functional)

$$J(x, u) = \boldsymbol{E}\Big(rX_N^2 + \sum_{k=1}^{N-1}(pX_k^2 + qu_k^2)\Big), \tag{9.17}$$

and where $r \geq 0$, $p \geq 0$, and $q > 0$. The initial value $X_0 = x$ is fixed.

1. Find the optimal control.