Lecture note 38. Controlled diffusion processes.

Sometimes we face a situation where not only we observe some process, but are able to influence how it goes: to control it. In these situations, problems may arise about how to control the process in the best way possible: problems of *optimal* control. In some of these situations no chance, no randomness is involved; in some other the process we are trying to control is a stochastic one. We are going to consider controlled *diffusion* processes.

For simplicity's sake, we are going to restrict our considerations to the case of time homogeneity.

There are several different models of controlled diffusion processes; we are going to consider one such model. Let us introduce the definitions.

Given functions $\boldsymbol{b}(\boldsymbol{x}) = \big(b_1(\boldsymbol{x}), ..., b_r(\boldsymbol{x})\big)$ and $\sigma(\boldsymbol{x}) = \big(\sigma_{ik}(\boldsymbol{x})\big)_{\substack{i=1,\,...,\,r\\ k=1,\,...,\,n}}$, $\boldsymbol{x} \in \mathbb{R}^r$, a diffusion process is a solution of a system of stochastic equations (a vector stochastic equation)

$$dX_t^i = b_i(\boldsymbol{X}_t)\,dt + \sum_{k=1}^{n} \sigma_{ik}(\boldsymbol{X}_t)\,dW_t^k, \quad i = 1, ..., r, \qquad d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{X}_t)\,dt + \sigma(\boldsymbol{X}_t)\,d\boldsymbol{W}_t. \tag{38.1}$$

Now suppose we have a variable u, taking values in a set U; we'll call u the control variable. Suppose we have two functions of $\boldsymbol{x} \in \mathbb{R}^r$ and $u \in U$:

$$\boldsymbol{b}(\boldsymbol{x}, u) = \big(b_1(\boldsymbol{x}, u), ..., b_r(\boldsymbol{x}, u)\big), \qquad \sigma(\boldsymbol{x}, u) = \big(\sigma_{ik}(\boldsymbol{x}, u)\big)_{\substack{i=1,\,...,\,r\\ k=1,\,...,\,n}}. \tag{38.2}$$

Our *control strategy* is a functional $u(t;\, \boldsymbol{x}_s, s \le t)$ of the past up to time t, taking values in U. The controlled process under this control strategy is, by definition, a solution $\boldsymbol{X}_t^u$ of the stochastic equation

$$d\boldsymbol{X}_t^u = \boldsymbol{b}\big(\boldsymbol{X}_t^u,\, u(t;\, \boldsymbol{X}_s^u, s \le t)\big)\,dt + \sigma\big(\boldsymbol{X}_t^u,\, u(t;\, \boldsymbol{X}_s^u, s \le t)\big)\,d\boldsymbol{W}_t. \tag{38.3}$$

I would like to mention that I have the same *letter* to denote just the variable u taking values in the set U, and for a *functional* $u(t;\, \boldsymbol{x}_s,\, s \le t)$ representing a strategy. However I cannot do anything about this – short of, perhaps, using a Gothic letter for the functional; but this won't do. Try not to get confused. The *letter* being the same, the *notations* are different, though (just a letter versus the same letter and a whole lot of stuff in parentheses after it).

Of course, for some control strategies a solution $\boldsymbol{X}_t^u$ may exist, while for some other it may not. The controller should not choose a control strategy for which no solution exists – otherwise no control is possible.

We could consider a generalization in which the set of possible controls u at a point $\boldsymbol{x} \in \mathbb{R}^r$ is not the same U, but depends on $\boldsymbol{x}$: $u \in U(\boldsymbol{x})$; but better we restrict ourselves to the simplest scheme.

It is natural to call a strategy a *Markov* one if it depends only on the value $\boldsymbol{x}_t$ of the function at the last moment observed. For such strategies, the controlled process $\boldsymbol{X}_t^u$ is a Markov one (a diffusion process).

We haven't spoken yet about *optimal control*; and this is because we haven't yet said anything about our goals in controlling: about our gain (or loss) under a control strategy $u = u(t;\, \boldsymbol{x}_s,\, s \le t)$.

Suppose we have a region $G \subseteq \mathbb{R}^r$; a function $g(\boldsymbol{x},\, u)$ is given for $\boldsymbol{x} \in G$, $u \in U$ (interpreted as our gain while using the control u at the space point $\boldsymbol{x}$ per time unit), and a function $\varphi(\boldsymbol{x})$, $\boldsymbol{x} \in \partial G$: payoff to us at the time τ_G^u of leaving the region G.

Our total gain under the strategy $u(t;\, \boldsymbol{x}_s,\, 0 \le s \le t)$ is defined as

$$\mathrm{Gain}^u = \varphi(\boldsymbol{X}^u_{\tau_G^u}) + \int_0^{\tau_G} g\big(\boldsymbol{X}^u_t,\, u(s;\, \boldsymbol{X}^u_s,\, 0 \le s \le t)\big)\, dt. \tag{38.4}$$

`[In the lecture, I wrote` $\varphi - \int$`, which contradicted some other formulas (I was confused with all these minuses); sorry, let it be` $+$`.]`

Note that our gain, and every part of it, can be negative, so it could be rather a loss; but this does not change the problem mathematically.

The time has come to make our notations more complicated by indicating the initial point: let $\boldsymbol{X}^{\boldsymbol{x},\, u}_t$ be the solution of the stochastic equation (38.3) with the initial condition $\boldsymbol{X}^u_0 = \boldsymbol{x}$; the stopping time

$$\tau_G^{\boldsymbol{x},\, u} = \min\{t \ge 0\colon\ \boldsymbol{X}^{\boldsymbol{x},\, u}_t \notin G\} \tag{38.5}$$

(or ∞ if the process does not leave the region at all), the total gain

$$\mathrm{Gain}^{\boldsymbol{x},\, u} = \varphi(\boldsymbol{X}^{\boldsymbol{x},\, u}_{\tau_G^{\boldsymbol{x},\, u}}) + \int_0^{\tau_G^{\boldsymbol{x},\, u}} g\big(\boldsymbol{X}^{\boldsymbol{x},\, u}_t,\, u(s;\, \boldsymbol{X}^{\boldsymbol{x},\, u}_s,\, 0 \le s \le t)\big)\, dt. \tag{38.6}$$

We want to maximize the expected gain

$$E(\mathrm{Gain}^{\boldsymbol{x},\, u}) = \max. \tag{38.7}$$

Why do we consider the cost function $g(\boldsymbol{x}, u)$, $\boldsymbol{x} \in G$, depending on the control variable u, and the boundary function $\varphi(\boldsymbol{x})$ *not* depending on it? It turns out that introducing the boundary gain function $\varphi(\boldsymbol{x}, u)$ depending also on the control does not lead to anything new and interesting. Suppose that for every $\boldsymbol{x} \in \partial G$ the maximum of the boundary gain function over all $u \in U$ is reached at some point $u_{\boldsymbol{x}} \in U$:

$$\tilde\varphi(\boldsymbol{x}) = \max_{u \in U} \varphi(x, u) = \varphi(\boldsymbol{x},\, u_{\boldsymbol{x}}) \ge \varphi(\boldsymbol{x},\, u) \qquad \text{for all } u \in U. \tag{38.8}$$

It is the first time that we introduce this requirement that the maximum of some function *is reached* at some point. Without this restriction we would have to consider a *supremum*, and our equalities would hold not precisely, but *up to an arbitrarily small* ε. To save us all this bother, we are going to assume in this case and in similar cases later that a minimum does exist.

It turns out that we can replace every control strategy $u(t;\, \boldsymbol{x}_s,\, 0 \le s \le t)$ with some other, $\tilde u(t;\, \boldsymbol{x}_s,\, 0 \le s \le t)$, that is *not worse than* $u(t;\, \boldsymbol{x}_s,\, 0 \le s \le t)$, and under this new strategy the total gain with the payoff function $\varphi(\boldsymbol{x}, u)$ is the same as that with the payoff function $\tilde{\varphi}(\boldsymbol{x})$ depending on the point $\boldsymbol{x} \in \partial G$ only; so that the problem of optimization with an arbitrary payoff function $\varphi(\boldsymbol{x},\, u)$ is reduced to one with the payoff function $\tilde{\varphi}(\boldsymbol{x})$.

As this control strategy $\tilde{u}$ we take the strategy u, changed only at the time

$$\tau = \min\{t\colon\, \boldsymbol{x}_t \notin G\} \tag{38.9}$$

at which the function x_t leaves the region G. We take

$$\tilde{u}(t;\, x_s,\, 0 \le s \le t) = \begin{cases} u(t;\, x_s,\, 0 \le s \le t) & \text{for } t \ne \tau, \\ \quad u_{\boldsymbol{x}_\tau} & \text{for } t = \tau. \end{cases} \tag{38.10}$$

The strategy $\tilde{u}$ yields a better (not a worse) result in gain than u:

$$\begin{aligned} &\varphi\big(\boldsymbol{X}^{\tilde{u}}_{\tau_G},\, \tilde{u}(\tau_G;\, \boldsymbol{X}^{\tilde{u}}_s,\, s \le \tau_G)\big) + \int_0^{\tau_G} g\big(\boldsymbol{X}^{\tilde{u}}_t,\, \tilde{u}(t;\, \boldsymbol{X}^{\tilde{u}}_s,\, s \le t)\big)\, dt \\ &\qquad \ge \varphi\big(\boldsymbol{X}^{u}_{\tau_G},\, u(\tau_G;\, \boldsymbol{X}^{u}_s,\, s \le \tau_G)\big) + \int_0^{\tau_G} g\big(\boldsymbol{X}^{u}_t,\, u(t;\, \boldsymbol{X}^{u}_s,\, s \le t)\big)\, dt. \end{aligned} \tag{38.11}$$

Indeed, before the time τ_G both control functions coincide, and so do the controlled processes $\boldsymbol{X}^{\tilde{u}}_t$ and $\boldsymbol{X}^{u}_t$ before the time τ_G and *at* this time; and the times at which $\boldsymbol{X}^{\tilde{u}}_t$ and $\boldsymbol{X}^{u}_t$ leave the region G are the same. So the integrals coincide; and $\varphi\big(\boldsymbol{X}^{\tilde{u}}_{\tau_G},\, \tilde{u}(\tau_G;\, \boldsymbol{X}^{\tilde{u}}_s,\, s \le \tau_G)\big) = \tilde{\varphi}(\boldsymbol{X}^{\tilde{u}}_{\tau_G}) \ge \varphi\big(\boldsymbol{X}^{u}_{\tau_G},\, u(\tau_G;\, \boldsymbol{X}^{u}_s,\, s \le \tau_G)\big)$ by (38.10).

There are other settings of optimization problems similar to (38.7). For example, it is widely acknowledged in the financial world that one dollar a year from now is worth less for us than one dollar today. To accommodate for this, we can consider, instead of (38.4), the gain functional

$$e^{-\beta\tau_G}\varphi(\boldsymbol{X}^u_{\tau_G}) + \int_0^{\tau_G} e^{-\beta t} g\big(\boldsymbol{X}^u_t,\, u(s;\, \boldsymbol{X}^u_s,\, 0 \le s \le t)\big)\, dt, \tag{38.12}$$

where β is some positive constant, and try to maximize its expectation.

Let us look what we can obtain in case of $\beta = 0$, which seems to be a little simpler (involving a little shorter formulas).

Let the diffusion coefficients under a control $u \in U$ be defined as

$$a_{ij}(\boldsymbol{x},\, u) = \sum_{k=1}^{n} \sigma_{ik}(\boldsymbol{x},\, u) \cdot \sigma_{jk}(\boldsymbol{x},\, u) \tag{38.13}$$

(in the matrix form: $a(\boldsymbol{x},\, u) = \sigma(\boldsymbol{x},\, u)\, \sigma(\boldsymbol{x},\, u)^T$). Let the linear differential operator L_u acting on twice continuously differentiable functions in G be defined by

$$L_u f(\boldsymbol{x}) = \frac{1}{2} \sum_{i,\, j=1}^{r} a_{ij}(\boldsymbol{x},\, u) \cdot \frac{\partial^2 f}{\partial x^i \partial x^j}(\boldsymbol{x}) + \sum_{i=1}^{r} b_i(\boldsymbol{x},\, u) \cdot \frac{\partial f}{\partial x^i}(\boldsymbol{x}). \tag{38.14}$$

Let us introduce the following differential operator:

$$\mathfrak{L} f(\boldsymbol{x}) = \max_{u \in U} \bigl(L_u f(\boldsymbol{x}) + g(\boldsymbol{x},\, u) \bigr) \tag{38.15}$$

(we suppose that the maximum *is* reached; otherwise we would have to consider the supremum, and our formulations would be much more complicated). For every $u \in U$ the operator $L_u + g(\boldsymbol{x},\, u)$ is "linear inhomogeneous": a linear operator, plus a function ($g(\boldsymbol{x},\, u)$) that does not depend on the function f. But the operator $\mathfrak{L}$ is non-linear, and even not "linear inhomogeneous".

The same we can see in a much more tangible case of *functions* instead of operators: the functions $f_a(x) = ax - a^2$ are definitely linear; while their maximum

$$\mathfrak{f}(x) = \max_{a \in \mathbb{R}} f_a(x) = \frac{x^2}{4} \tag{38.16}$$

is definitely nonlinear. (Draw a picture of the graphs of $f_a(x)$ for $a = 0,\ \pm 1/2,\ \pm 1,\ \pm 3/2,\ \pm 2,$ and of $\mathfrak{f}(x)$.)

Theorem 38.1. *Let G be a region in $\mathbb{R}^r$. Suppose a function $v(\boldsymbol{x})$, $\boldsymbol{x} \in G \cup \partial G$ is a solution of the nonlinear Dirichlet problem*

$$\begin{aligned} \mathfrak{L} v(\boldsymbol{x}) &= 0, \qquad \boldsymbol{x} \in G, \\ v(\boldsymbol{x}) &= \varphi(\boldsymbol{x}), \quad \boldsymbol{x} \in \partial G, \end{aligned} \tag{38.17}$$

and suppose that this function can be extended to the whole $\mathbb{R}^r$ to form a function that is twice continuously differentiable and bounded together with its first and second derivatives.

Let $\hat{u}(\boldsymbol{x})$ be the value (or one of such values) *at which $\max_{u \in U} L_u v(\boldsymbol{x})$ is reached*: *$\mathfrak{L} v(\boldsymbol{x}) = L_{\hat{u}(\boldsymbol{x})} v(\boldsymbol{x})$ for $\boldsymbol{x} \in G$; and let $\hat{u}(\boldsymbol{x})$ be defined arbitrarily for $\boldsymbol{x} \in \mathbb{R}^r \setminus G$. Let us consider the control strategy*

$$\hat{u}(t;\, \boldsymbol{x}_s,\, 0 \le s \le t) = \hat{u}(\boldsymbol{x}_t) \tag{38.18}$$

depending only on the value of the function at the last observed time. Suppose the stochastic equation

$$d\boldsymbol{X}_t^{\hat{u}} = \boldsymbol{b}\bigl(\boldsymbol{X}_t^{\hat{u}},\, \hat{u}(\boldsymbol{X}_t^{\hat{u}})\bigr)\, dt + \sigma\bigl(\boldsymbol{X}_t^{\hat{u}},\, \hat{u}(\boldsymbol{X}_t^{\hat{u}})\bigr)\, d\boldsymbol{W}_t \tag{38.19}$$

has a solution for every initial value $\boldsymbol{X}_0^{\hat{u}} = \boldsymbol{x} \in G$, and $E(\tau_G^{\boldsymbol{x},\hat{u}}) < \infty$.

Then the control strategy (38.18) *maximizes the expectation* (38.7) *among all control strategies $u(t;\, \boldsymbol{x}_s,\, 0 \le s \le t)$ for which $E(\tau_G^{\boldsymbol{x},\, u}) < \infty$ and the function*(al) *$g\bigl(\boldsymbol{x}_t,\, u(t;\, \boldsymbol{x}_s,\, 0 \le s \le t)\bigr)$ is bounded.*

So we always have a Markov optimal strategy (there could be also non-Markov strategies that are just as good; but not *better*).

Another thing I want to attract your attention to is that the maximum is reached at the same strategy $\hat{u}$ for our expectations for *all* starting points $\boldsymbol{x}$: no situation of us using one strategy for one starting point, and a different one for another.

The nonlinear equation $\mathfrak{L} v(\boldsymbol{x}) = 0$ and other similar equations are associated with the name of Bellman, an American mathematician: Bellman's equation.

The proof is relegated to another lecture note.