12. Optimal Controls and Dynamic Programming.

One of the useful concepts in optimization is the notion of Dynamic Programming. Let us suppose that we have a collection of trnasition probability densities on Z, $\{p(\alpha, i, j)\}$ that depend on a parameter α that varies over a set $\mathcal{A}$. At each step we can choose α as we please and if we make the choice of α_n at t time n while in state i, then the transition to the state j at time $n+1$ takes palce with probability $p(\alpha, i, j)$. We do not have to make the choice of α before time n, so that α_n can be sa function $\alpha_n(X_0, X_1, \ldots, X_n)$ of the history through time n. After making a final choice $\alpha_{N_1}(X_0, X_1, \ldots, X_{N_1})$ we arrive at some random state X_N and the game ends. The goal is to make the choices $\{\alpha_j : 0 \leq j \leq N-1\}$ in order to maximize the expected payoff $E[f(X_N)|X_0 = i]$. It is clear that if we maximize all that we can there is an optimal value, which is actually attained under mild regularity conditions. If we start from x at time k and play till the end there is an optimal value that we call $V(k, i)$. The principle of dynamic programming allows us to calculate $V(k, i)$ recursively. Cleacrly $V(N, i) = f(i)$ since the game is over. We can compute $V(N-1, i)$ because if we make the choice of α the expected payoff is

$$\sum_j f(j)p(\alpha, i, j)$$

and it is clearly the right strategy to maximize the above expression and

$$V(N-1, i) = \sup_{\alpha \in \mathcal{A}} \sum_j f(j)p(\alpha, i, j) = \sup_{\alpha \in \mathcal{A}} \sum_j V(N, j)p(\alpha, i, j)$$

Since we know that the best we can do at time $N-1$ is to choose $\alpha_{N-1} = \alpha(N-1, i)$ that yielsds the maximum in the step above, we can pretend that the game ends at time $N-1$ with a payoff of $V(N-1, j)$. Now we can repeat the process to define inductively

$$V(k, i) = \sup_{\alpha \in \mathcal{A}} \sum_j V(k+1, j)p(\alpha, i, j)$$

and $\alpha(k, i)$ is the α that achieves the maximum. We stop when we reach $V(0, i)$ which is the optimal value and the optimal choice of α at time k depends only on k and the current state X_k and is $\alpha(k, X_k)$. the computational algorithm yields successively $V(k, i)$ and $\alpha(k, i)$. It is clear that the choice of

$\{\alpha(k, X_k)\}$ will produce the expected payoff of $V(0, i)$ if $X_0 = i$. Let us show that this cannot be improved. Let $\{\alpha(k, X_0, X_1, \ldots, X_k) : 0 \leq k \leq N-1\}$ be an arbitrary sequence of choices. Then

$$\begin{aligned} & E\Big[V(k+1, X_{k+1})\Big|X_0, X_1, \ldots, X_k\Big] \\ &= \sum_j V(k, j)p(\alpha(k, X_0, X_1, \ldots, X_k), X_k, j) \\ &\leq \sup_{\alpha \in \mathcal{A}} \sum_j V(k, j)p(\alpha, X_k, j) \\ &= V(k, X_k) \end{aligned}$$

proceeding inductively we conclude that

$$E\Big[f(X_N)\Big|X_0\Big] \leq V(0, X_0)$$

The strategy provided by the dynamic programming principle cannot be improved. We can consider instead expected discounted payoff over time. Try to optimize

$$\sup_{\{\alpha(\cdot)\}} E\left[\sum_{r=0}^{\infty} \rho^r f(X_r)|X_0 = i\right]$$

If the optimal value is $U(i)$, then clearly

$$U(x) = f(i) + \rho \sup_{\alpha} \sum_j U(j)p(\alpha, i, j)$$

For $\rho < 1$ this equation has a unique sloution and the value $\alpha(i)$ of α that maximizes provides the strategy. It does not explicitly depend on n because the time horizon is infinite. We have used the argument that if the optimal expected total discounted payoff is $U(i)$ when we start from i, if we find ourselves in $X_1 = j$ at time 1, the best we can do is $\rho U(j)$ or $U(X_1)$. Therefore the best α initially is one that maximizes

$$\rho E\left[U(X_1)|X_0 = i\right] = \rho \sum_j U(j)p(\alpha, i, j)$$

which gives us the equation for U.

Now let us move to continuous time and SDE's. Let us consider a stochastic differential equation

$$dx(t) = \sigma(t, x(t), u(t))d\beta(t) + b(t, x(t), u(t))dt \; ; \qquad x(0) = x_0 \qquad (1)$$

where $u(t) = u(t, \omega)$ is something we can control on the basis of past information. Naturally, we want to make the choice of $u(t)$ with the aim of optimizing the expected value of some utility. The utility may be a combination of two terms. The first term is of the form

$$\int_0^T f(t, x(t), u(t))dt$$

and involves both the utility derived from $x(t)$ and the cost of exercising the control $u(t)$. The second term involves the terminal utility $g(x(T))$. We want to pick u in order to maximize

$$E\left[\int_0^T f(t, x(t), u(t))dt + g(x(T))|x(0) = x_0\right]$$

Of course if we pick the control to depend on past history and not just on the current state $x(t)$ we get out of the cass od Markov type SDEs that we considerd till now. But we will see that such a choice will not be needed. It is convenient to consider the 'value function'

$$V(s, x) = \sup_{u \in \mathcal{U}} E\left[\int_s^T f(t, x(t), u(t))dt + g(x(T))|x(s) = x\right]$$

over all possible admissible contorls. $\mathcal{U}$ consisits of functions $u(t, \omega)$ that are reasonable functions of the past history such that $u(t, \omega) \in \mathbf{U}$ for every t. here the set $\mathbf{U}$ is the set of allowed values of the control parameter u. Our goal is to derive a PDE satisfied by $V(t, x)$ and determine the optimal choice of $u(\cdot, \cdot)$ that will realize it. Let us fist limit overselves to Markoivan controls $u(t, \omega) = u(t, x(t))$. If at time s we pick the control $u(s, x(s)) = u$ and hold it for time $h > 0$, and do the optimal thing after time $s + h$, the expected utility will be the solution of

$$W_t + \frac{1}{2}a(t, x, u))W_{xx} + b(t, x, u)W_x + f(t, x, u) = 0$$

$$W(s + h, x) = V(t + h, x)$$

giving an approximate value of

$$V(s+h,x)+h\left[\frac{1}{2}a(t,x,u)V_{xx}(s+h,x)+b(t,x,u)V_x(s+h,x)+f(t,x,u)\right]$$

for $W(s,x)$ with an error of $o(h)$. Since we can optimize over u,

$$\begin{aligned}&\frac{V(s,x)-V(s+h,x)}{h}\\&\simeq\sup_{u\in\mathbf{U}}\left[\frac{1}{2}a(s,x,u)V_{xx}(s+h,x)+b(s,x,u)V_x(s+h,x)+f(s,x,u)\right]\end{aligned}$$

or

$$V_t+\sup_{u\in\mathbf{U}}\left[\frac{1}{2}a(t,x,u)V_{xx}(t,x)+b(t,x,u)V_x(t,x)+f(t,x,u)\right]=0$$

with $V(T,x)=g(x)$. Note that this is a nonlinewr PDE for V

$$V_t+F(t,x,V_x,V_{xx})=0$$

with

$$F(t,x,p,q)=\sup_{u\in\mathbf{U}}\left[\frac{1}{2}a(t,x,u)q+b(t,x,u)p+f(t,x,u)\right]$$

Note also that the best choice of u is given by the maximising

$$u=u(t,x,p,q)=u(t,x,V_x(t,x),V_{xx}(t,x)).$$

In particular with this choice the utility $V(s,x)$ is realized. Suppose $u(t,\omega)$ is **any** admissible choice of the control, then

$$\begin{aligned}&\psi(t,\omega)\\&=V_t+\left[\frac{1}{2}a(t,x,u(t,\omega))V_{xx}(t,x)+b(t,x,u(t,\omega))V_x(t,x)+f(t,x,u(t,\omega))\right]\\&\leq 0\end{aligned}$$

From Itô's formula, for the solution

$$dx(t)=\sigma(t,x(t),u(t,\omega)d\beta(t)+b(t,x(t),u(t,\omega)dt$$

$$V(s,x) = E\left[\int_s^T [f(t,x(t),u(t,\omega)) - \psi(t,\omega)]dt + g(x(T))\right]$$
$$\geq E\left[\int_s^T f(t,x(t),u(t,\omega))dt + g(x(T))\right]$$

proving the optimality of $u(t,x,V_x(t,x),V_{xx}(t,x))$ and $V(s,x)$ as the optimal value.

If we want to minimize the cost rather than maximize the utility the **sup** turns into an **inf**.

Examples.

1. Let us consider the problem of minimizing

$$E\left[x^2(T) + c\int_s^T u^2(t)dt | x(s) = x\right]$$

withsome $c > 0$, for solutions of

$$dx(t) = d\beta(t) + u(t,\omega)dt$$

Here, $f = 0$ and $g(x) = x^2$.

$$F(t,x,p.q) = \inf_{-\infty<u<\infty}\left[\frac{1}{2}q + pu + cu^2\right] = \frac{q}{2} - \frac{p^2}{4c}$$

So the equation to solve is

$$V_t + \frac{1}{2}V_{xx} - \frac{V_x^2}{4c} = 0$$

with $V(T,x) = x^2$. If we try

$$V(t,x) = a(t)x^2 + b(t)$$

we get

$$a'(t)x^2 + b'(t) + a(t) - \frac{[a(t)]^2x^2}{c}$$

Equating coefficients of 1 and x^2,

$$a'(t) = \frac{[a(t)]^2}{c}$$
$$b'(t) = -a(t)$$

with $a(T) = 1$ and $b(T) = 0$. Solving we get

$$a(t) = \frac{c}{c + T - t}$$
$$b(t) = c \log(1 + \frac{T - t}{c})$$

providing us with the value

$$V(s, x) = \frac{cx^2}{c + T - t} + c \log(1 + \frac{T - t}{c})$$

and

$$u(t, x) = \frac{-V_x}{2c} = -\frac{a(t)x}{c} = -\frac{x}{c + T - t}$$

tells us that the best control is

$$u(t, \omega) = \frac{-x(t)}{c + T - t}$$

2. Infinite time horizon problems. Consider a control problem with a generator $\mathcal{L}_u$ that depnds on a control parameter u. The utility is the discounted functional

$$E\left[\int_0^\infty f(x(t), u(t))e^{-\lambda t}dt | x(0) = x\right]$$

With out control we will solve

$$\lambda V - \mathcal{L}V = f$$

But with control, we now solve

$$\lambda V - \sup_u [(\mathcal{L}_u V)(x) + f(x, u)] = 0$$

Rest of the proof proceeds as before.

3. Problems with fixed boundary. Consider the problem of minimzing for a constant $c > 0$,

$$E\left[c^2\tau + \int_0^\tau u^2(t)dt \,|x(0) = x\right]$$

for the controlled process

$$dx(t) = d\beta(t) + u(t,\omega)dt$$

with τ as the first exit time from $[-1, 1]$.

The equation to solve is

$$\frac{1}{2}V_{xx} + \inf_u \left[uV_x + c^2 + u^2\right] = 0$$

with $V(\pm 1) = 0$.

$$\frac{V_{xx}}{2} - \frac{V_x^2}{4} = -c^2$$

Let $V_x = f(x)$. Then $f'(x) = -2c^2 + \frac{[f(x)]^2}{2}$. It can be integrated to give

$$V(x) = 2\log\frac{\cosh c}{\cosh cx}$$

4. Smoothness may be a problem. Consider the problem of minimzing f

$$E\left[\tau \,|x(0) = x\right]$$

for the controlled process

$$dx(t) = d\beta(t) + u(t,\omega)dt$$

with τ as the first exit time from $[-1, 1]$. But we limit the control u to the inrterval $[-A, A]$. The equation to solve is

$$\frac{1}{2}V_{xx} + AV_x = 1$$

for $x > 0$ with $V(1) = V'(0) = 0$ and define $V(x) = V(-x)$ to be even for $x < 0$. This is OK provided $V_x < 0$ for $x > 0$. We can solve it. But why does this work? With the sign right the inf is OK. Linear optimization leads

to extreme points as optimizers. The solution is not twice differentiable at the origin. There is a discontinuity of the second derivative. But the first derivative is continuous. This is enough for applying Itô's formula.

5*. Consider the problem of minimizing

$$E\left[\int_0^\infty [x^2(t) + c\ u^2(t)]e^{-\lambda t}dt | x(0) = x\right]$$

with some $c > 0$, for solutions of

$$dx(t) = d\beta(t) + u(t,\omega)dt$$

Show that there is an optimal control of the form $u(t,\omega) = u(x(t))$ that is time homogeneous. Find the minimum value $V(x)$.

6*. If we want to minimize

$$E\left[\int_0^\infty |x(t)|e^{-t}dt \,|\, x(0) = x\right]$$

where

$$dx(t) = d\beta(t) + u(t,\omega)dt$$

and the values of u are limited to the set $\mathbf{U} = [-A, A]$, what equation will you solve and how will you proceed? What is the solution?