12 Dynamic Programming in Continuous Time

We develop the HJB equation for dynamic programming in continuous time.

12.1 The optimality equation

In continuous time the plant equation is,

$$\dot{x} = a(x, u, t) \,.$$

Let us consider a discounted cost of

$$\mathbf{C} = \int_0^T e^{-\alpha t} c(x, u, t) \, dt + e^{-\alpha T} \mathbf{C}(x(T), T) \, .$$

The discount factor over δ is $e^{-\alpha\delta} = 1 - \alpha\delta + o(\delta)$. So the optimality equation is,

$$F(x,t) = \inf_{\alpha} \left[c(x,u,t)\delta + e^{-\alpha\delta}F(x+a(x,u,t)\delta,t+\delta) + o(\delta) \right] .$$

By considering the term that multiplies δ in the Taylor series expansion we obtain,

$$\inf_{u} \left[c(x, u, t) - \alpha F + \frac{\partial F}{\partial t} + \frac{\partial F}{\partial x} a(x, u, t) \right] = 0, \quad t < T,$$
(12.1)

with $F(x,T) = \mathbf{C}(x,T)$. In the undiscounted case, we simply put $\alpha = 0$.

The DP equation (12.1) is called the **Hamilton Jacobi Bellman equation** (HJB). Its heuristic derivation we have given above is justified by the following theorem.

Theorem 12.1 Suppose a policy π , using a control u, has a value function F which satisfies the HJB equation (12.1) for all values of x and t. Then π is optimal.

Proof. Consider any other policy, using control v, say. Then along the trajectory defined by $\dot{x} = a(x, v, t)$ we have

$$-\frac{d}{dt}e^{-\alpha t}F(x,t) = e^{-\alpha t}\left[c(x,v,t) - \left(c(x,v,t) - \alpha F + \frac{\partial F}{\partial t} + \frac{\partial F}{\partial x}a(x,v,t)\right)\right]$$
$$\leq e^{-\alpha t}c(x,v,t).$$

Integrating this inequality along the v path, from x(0) to x(T), gives

$$F(x(0), 0) - e^{-\alpha T} \mathbf{C}(x(T), T) \le \int_{t=0}^{T} e^{-\alpha t} c(x, v, t) dt$$

Thus the v path incurs a cost of at least F(x(0), 0), and hence π is optimal.

12.2 Example: LQ regulation

The undiscounted continuous time DP equation for the LQ regulation problem is

$$0 = \inf_{u} \left[x^{\top} R x + u^{\top} Q u + F_t + F_x^{\top} (A x + B u) \right] \,.$$

Suppose we try a solution of the form $F(x,t) = x^{\top} \Pi(t)x$, where $\Pi(t)$ is a symmetric matrix. Then $F_x = 2\Pi(t)x$ and the optimizing u is $u = -\frac{1}{2}Q^{-1}B^{\top}F_x = -Q^{-1}B^{\top}\Pi(t)x$. Therefore the DP equation is satisfied with this u if

$$0 = x^{\top} \left[R + \Pi A + A^{\top} \Pi - \Pi B Q^{-1} B^{\top} \Pi + \frac{d\Pi}{dt} \right] x$$

where we use the fact that $2x^{\top}\Pi Ax = x^{\top}\Pi Ax + x^{\top}A^{\top}\Pi x$. Hence we have a solution to the HJB equation if $\Pi(t)$ satisfies the Riccati differential equation of Section 7.4.

12.3 Example: estate planning

A man is considering his lifetime plan of investment and expenditure. He has an initial level of savings x(0) and no other income other than that which he obtains from investment at a fixed interest rate. His total capital is therefore governed by the equation

$$\dot{x}(t) = \beta x(t) - u(t) \,,$$

where $\beta > 0$ and u is his rate of expenditure. He wishes to maximize

$$\int_0^T e^{-\alpha t} \sqrt{u(t)} \, dt,$$

for a given T. Find his optimal policy.

Solution. The optimality equation is

$$0 = \sup_{u} \left[\sqrt{u} - \alpha F + \frac{\partial F}{\partial t} + \frac{\partial F}{\partial x} (\beta x - u) \right] \,.$$

Suppose we try a solution of the form $F(x,t) = f(t)\sqrt{x}$. For this to work we need

$$0 = \sup_{u} \left[\sqrt{u} - \alpha f \sqrt{x} + f' \sqrt{x} + \frac{f}{2\sqrt{x}} (\beta x - u) \right].$$

By $d[\]/du = 0$, the optimizing u is $u = x/f^2$ and the optimized value is

$$(\sqrt{x}/f)\left[\frac{1}{2} - (\alpha - \frac{1}{2}\beta)f^2 + ff'\right].$$
 (12.2)

We have a solution if we can choose f to make the bracketed term in (12.2) equal to 0. We have the boundary condition F(x, T) = 0, which imposes f(T) = 0. Thus we find

$$f(t)^{2} = \frac{1 - e^{-(2\alpha - \beta)(T - t)}}{2\alpha - \beta}$$

We have found a policy whose value function F(x,t) satisfies the HJB equation. So by Theorem 12.1 it is optimal. In closed loop form the optimal policy is $u = x/f^2$.

12.4 Example: harvesting

A fish population of size x obeys the plant equation,

$$\dot{x} = a(x, u) = \begin{cases} a(x) - u & x > 0, \\ a(x) & x = 0. \end{cases}$$

The function a(x) reflects the facts that the population can grow when it is small, but is subject to environmental limitations when it is large. It is desired to maximize the discounted total harvest $\int_0^T u e^{-\alpha t} dt$.

Solution. The DP equation (with discounting) is

$$\sup_{u} \left[u - \alpha F + \frac{\partial F}{\partial t} + \frac{\partial F}{\partial x} [a(x) - u] \right] = 0, \quad t < T$$

Hence u occurs linearly with the maximization and so we have a bang-bang optimal control of the form

$$u = \begin{bmatrix} 0 \\ \text{undetermined} \\ u_{\text{max}} \end{bmatrix} \text{ for } F_x \begin{bmatrix} > \\ = \\ < \end{bmatrix} 1,$$

where u_{max} is the largest practicable fishing rate.

Suppose $F(x,t) \to F(x)$ as $T \to \infty$, and $\partial F/\partial t \to 0$. Then

$$\sup_{u} \left[u - \alpha F + \frac{\partial F}{\partial x} [a(x) - u] \right] = 0.$$
 (12.3)

Let us make a guess that F(x) is concave, and then deduce that

$$u = \begin{bmatrix} 0 \\ \text{undetermined, but effectively } a(\bar{x}) \\ u_{\max} \end{bmatrix} \text{ for } x \begin{bmatrix} < \\ = \\ > \end{bmatrix} \bar{x}.$$
(12.4)

Clearly, \bar{x} is the operating point. We suppose

$$\dot{x} = \begin{cases} a(x) > 0, & x < \bar{x} \\ a(x) - u_{\max} < 0, & x > \bar{x}. \end{cases}$$

We say that there is **chattering** about the point \bar{x} , in the sense that u switches between its maximum and minimum values either side of \bar{x} , effectively taking the value $a(\bar{x})$ at \bar{x} . To determine \bar{x} we note that

47

$$F(\bar{x}) = \int_0^\infty e^{-\alpha t} a(\bar{x}) dt = a(\bar{x})/\alpha \,. \tag{12.5}$$

So from (12.3) and (12.5) we have

$$F_x(x) = \frac{\alpha F(x) - u(x)}{a(x) - u(x)} \to 1 \text{ as } x \nearrow \bar{x} \text{ or } x \searrow \bar{x}.$$
(12.6)

For F to be concave, F_{xx} must be negative if it exists. So we must have

$$F_{xx} = \frac{\alpha F_x}{a(x) - u} - \left(\frac{\alpha F - u}{a(x) - u}\right) \left(\frac{a'(x)}{a(x) - u}\right)$$
$$= \left(\frac{\alpha F - u}{a(x) - u}\right) \left(\frac{\alpha - a'(x)}{a(x) - u}\right)$$
$$\simeq \frac{\alpha - a'(x)}{a(x) - u(x)}$$

where the last line follows because (12.6) holds in a neighbourhood of \bar{x} . It is required that F_{xx} be negative. But the denominator changes sign at \bar{x} , so the numerator must do so also, and therefore we must have $a'(\bar{x}) = \alpha$. Choosing this as our \bar{x} , we have that F(x) is concave, as we conjectured from the start.

We now have the complete solution. The control in (12.4) has a value function F which satisfies the HJB equation.



Growth rate a(x) subject to environment pressures

Notice that there is a sacrifice of long term yield for immediate return. If the initial population is greater than \bar{x} then the optimal policy is to overfish at u_{\max} until we reach the new \bar{x} and then fish at rate $u = a(\bar{x})$. As $\alpha \nearrow a'(0)$, $\bar{x} \searrow 0$. So for sufficiently large α it is optimal to wipe out the fish population.