INTRODUCTION

The study of dynamical systems and the consequent mathematical theories have been making highly significant contributions in virtually every one of scientific branches. This course is concerned with is the control theory and an extremely important and elegant viewpoint introduced by Richard Bellman in 1950's, namely *dynamic programming* (DP). Some references are listed in the end of this chapter [2, 3, 4, 5, 6, 7, 8, 9, 10]

Even though *stochastic* control theory is the main subject for this course, the idea of DP can be fully illustrated by simple *deterministic* models. We should do so, and introduce some terminologies at the same time, through an example of finite-time-horizon utility maximization problem (in both discrete time and continuous time).

Example 1 (Discrete-time utility maximization). Consider an agent with initial wealth X_0. At each time point $n \in \{0, 1, \ldots, N-1\}$, the agent determines to consume a proportion $u_n \in [0,1]$ of the total available wealth X_n, and put the rest in the bank. The bank will offer an interest rate of $r > 0$ per unit of time. Write $R = 1 + r$, then the *dynamics* of the total wealth process is governed by

$$X_{n+1} = RX_n(1-u_n); \qquad n = 0, \ldots, N-1. \tag{1}$$

For the amount of wealth u_nX_n consumed at time n, the agent will receive a utility of $f(u_nX_n)$. For the amount of wealth X_N left at the *terminal* time N, the agent will receive a utility of $g(X_N)$. The optimization problem for the agent is to come up with a sequence of consumption $\{u_n : n = 0, \ldots, N-1\}$ so as to maximize the total utility

$$\sum_{n=0}^{N-1} f(u_nX_n) + g(X_N).$$

How should the agent proceed to find this optimal consumption sequence?

The solution is divided into two steps. The first step is to find a *candidate* optimal solution via DP. The second step is to *verify* that the candidate solution from Step 1 is indeed the true optimal solution.

From now on, we will denote the maximum of total utility by

$$v(X_0) \doteq \max_{\{u_n\}} \left[\sum_{n=0}^{N-1} f(u_nX_n) + g(X_N) \right]. \tag{2}$$

Step 1: The idea of DP, roughly speaking, is very much like *backward* induction. The purpose is to divide the entire, complicated dynamic problem into small, solvable, static sub-problems.

Suppose the agent is at the last time period $n = N - 1$ with a total wealth, say x. Clearly the agent should consume the proportion u_{N-1} which maximize the total utility

$$f(u_{N-1}x) + g\left(R(1 - u_{N-1})x\right).$$

If we denote the maximal value by $V_{N-1}(x)$, then

$$V_{N-1}(x) = \sup_{0 \leq u_{N-1} \leq 1} \left[f(u_{N-1}x) + g\left(R(1 - u_{N-1})x\right) \right].$$

This is a usual static calculus problem! Given the function f and g, *function* V_{N-1} can be calculated either analytically or numerically. The maximizer u^*_{N-1} can also be obtained, and it (usually) depends on the wealth x at time $n = N - 1$. Note that x is generic. The procedure is true for any $x \in \mathbb{R}_+$.

Suppose the agent is at the time period $n = N - 2$ with a total wealth, say again x. Clearly the agent should look for a consumption proportion u_{N-2} so as to maximize

$$f(u_{N-2}x) + V_{N-1}\left(R(1 - u_{N-2})x\right).$$

If we denote the maximal value by $V_{N-2}(x)$, then

$$V_{N-2}(x) = \sup_{0 \leq u_{N-2} \leq 1} \left[f(u_{N-2}x) + V_{N-1}\left(R(1 - u_{N-2})x\right) \right].$$

This again, is a static calculus problem, since we have already obtained V_{N-1}! The maximizer u^*_{N-2} is also dependent on x, the wealth at $n = N-2$. Again, x is generic and the procedure is valid for any $x \in \mathbb{R}_+$

Similarly, suppose the agent is at time n with total wealth x. The agent should solve the static optimization problem

$$V_n(x) = \sup_{0 \leq u_n \leq 1} \left[f(u_n x) + V_{n+1}\left(R(1 - u_n)x\right) \right].$$

Intuitively, $V_n(x)$ has the interpretation as the *maximal total utility possible from time n on if the total wealth at time n is x.*

In conclusion, if we let $V_N(x) \doteq g(x)$, and let

$$V_n(x) = \sup_{0 \leq u_n \leq 1} \left[f(u_n x) + V_{n+1}\left(R(1 - u_n)x\right) \right], \quad n = 0, 1, \ldots, N - 1, \tag{3}$$

then one would naturally expect $v(X_0) = V_0(X_0)$, and that the following sequentially defined consumption sequence $\{u_n^*\}$ is optimal:

$$\begin{aligned}
u_0^* &\doteq u_0^*(X_0) && (4)\\
X_1^* &= R(1-u_0^*)X_0 && (5)\\
u_1^* &\doteq u_1^*(X_1^*) && (6)\\
X_2^* &= R(1-u_1^*)X_1^* && (7)\\
&\vdots
\end{aligned}$$

Step 2: We will prove that $v(X_0) = V_0(X_0)$ and that the sequence $\{u_n^*\}$ defined by equations (4)-(7) is optimal. The proof goes as follows. First we show that for any consumption sequence $\{u_n\}$, the corresponding total utility never exceeds $V_0(X_0)$. Secondly we show that the total utility associated with consumption sequence $\{u_n^*\}$ is exactly $V_0(X_0)$. It follows then $v(X_0) = V_0(X_0)$ and $\{u_n^*\}$ is optimal.

Now let $\{u_n\}$ be an arbitrary consumption sequence, and $\{X_n\}$ the corresponding wealth process. By the definition (3) of $\{V_n\}$, we have

$$\begin{aligned}
V_0(X_0) &\geq V_1(R(1-u_0)X_0) + f(u_0X_0)\\
&= V_1(X_1) + f(u_0X_0)\\
&\geq V_2(R(1-u_2)X_1) + f(u_1X_1) + f(u_0X_0)\\
&= V_2(X_2) + f(u_1X_1) + f(u_0X_0)\\
&\vdots\\
&\geq V_N(X_N) + \sum_{n=0}^{N-1} f(u_nX_n)\\
&= g(X_N) + \sum_{n=0}^{N-1} f(u_nX_n).
\end{aligned}$$

But if we replace $\{u_n\}$ and $\{X_n\}$ by $\{u_n^*\}$ and $\{X_n^*\}$ respectively, all the inequalities become equalities. Or,

$$V_0(X_0) = g(X_N^*) + \sum_{n=0}^{N-1} f(u_n^*X_n^*).$$

It follows readily that $V_0(X_0) = v(X_0)$ and $\{u_n^*\}$ is an optimal consumption sequence. ■.

Remark **1** A small loop-hole in the above analysis is that we have implicitly assumed the existence of the optimizer $\{u_n^*(x)\}$ for equation (3). This assumption is not restrictive since virtually all the problems we will encounter have this property. Furthermore, even if this is not the case, the claim $v(X_0) = V_0(X_0)$ is still valid, though the existence of an optimal consumption sequence will be in question.

The equation (3) sometimes is said to be the *Dynamic Programming Equation* (DPE). Ideally, one would like to use the DPE to obtain (recursively) closed-form expressions of $\{V_n\}$. This is too much to ask in many practical cases, and one has to resort to numerical algorithms instead. However, most of the examples in this course will lead to at lease semi-explicit analytical solutions. Even though these examples have to use over-simplified assumption, they will provide valuable insights about the structure of the optimal solutions in more complex models.

Example 2 In the preceding example, set $f(x) = g(x) = x^\alpha$ for some $\alpha \in (0,1)$. This form of utility function is often called the "power utility" or "constant relative risk aversion utility". Solve explicitly $\{V_n\}$ and the optimal consumption sequence.

Starting with $n = N-1$, we have

$$\begin{aligned} V_{N-1}(x) &= \sup_{0 \le u_{N-1} \le 1} \left[f(u_{N-1}x) + g\left(R(1-u_{N-1})x\right) \right] \\ &= x^\alpha \cdot \sup_{0 \le u_{N-1} \le 1} \left[u_{N-1}^\alpha + R^\alpha (1-u_{N-1})^\alpha \right] \\ &\doteq c_{N-1} x^\alpha . \end{aligned}$$

The optimizing u_{N-1}^* is *independent* of x.

Taking $n = N-2$, we have

$$\begin{aligned} V_{N-2}(x) &= \sup_{0 \le u_{N-2} \le 1} \left[f(u_{N-2}x) + V\left(R(1-u_{N-2})x\right) \right] \\ &= x^\alpha \cdot \sup_{0 \le u_{N-2} \le 1} \left[u_{N-2}^\alpha + c_{N-1} R^\alpha (1-u_{N-2})^\alpha \right] \\ &\doteq c_{N-2} x^\alpha . \end{aligned}$$

Again the optimizing u_{N-2}^* is *independent* of x.

Repeating the above process, we have, for every $n = 0, 1, \ldots, N$,

$$V_n(x) = c_n x^\alpha .$$

The constants $\{c_n\}$ are recursively determined by

$$\begin{aligned} c_N &\doteq 1 \\ c_n &\doteq \sup_{0 \le u_n \le 1} \left[u_n^\alpha + c_{n+1} R^\alpha (1 - u_n)^\alpha \right], \quad n = 0, \ldots, N-1. \end{aligned} \tag{8}$$

The optimal consumption sequence $\{u_n^*\}$ is a sequence of fixed proportions, with each u_n^* the maximizer for the right-hand-side (RHS) of equation (8).

We conclude that, if the utility functions are of power type, the maximum utility function is also of power type, and the optimal consumption proportion may depend on the current time, but is independent of the current total wealth. ■

Terminologies: The function v is said to be the *value function*. The sequence $\{u_n\}$ is called the *control*. The process $\{X_n\}$ is the *controlled state process*, or simply the state process. The sequence $\{u_n^*\}$ is an *optimal control*. The terminal time N is said to the *time-horizon*. Equations like (1) is usually referred to as the *system dynamics*.

Remark **2** It is more precise to write the state process as $\{X_n^u\}$ in order to make clear the dependence of the system dynamics on the controls. However, to ease exposition, the index u is usually omitted when no confusion is incurred.

Our first theorem is concerned with general finite-time-horizon deterministic control problems.

Theorem 1 *Fix a time-horizon $N \ge 1$. Suppose the system dynamics are governed by*

$$X_{n+1} = h_n(X_n, u_n), \qquad n = 0, \ldots, N-1.$$

The sequence $\{u_n\}$ is the control. It is assumed that, for each n, the control $u_n \in U_n(X_n)$, where $U_n(X_n)$ is all available controls given the state X_n at time n. For an arbitrary initial state X_0, consider the optimization problem

$$v(X_0) \doteq \inf_{\{u_n\}} \left[g(X_N) + \sum_{n=0}^{N-1} f_n(X_n, u_n) \right].$$

Then the value function v equals the function V_0 that is recursively determined by the DPE

$$\begin{aligned} V_N(x) &\doteq g(x) \\ V_n(x) &\doteq \inf_{u_n \in U_n(x)} \left[f_n(x, u_n) + V_{n+1}(h_n(x, u_n)) \right], \quad 0, \ldots, N-1. \end{aligned}$$

Furthermore, if there exists a sequence $\{u_n^* \doteq u_n^*(x) \in U_n(x)\}$ *that attains the infimum of the RHS, then* $\{u_n^*\}$ *is an optimal control sequence.*

Proof. Analogous to Example 1, it is not difficult to show that, for any sequence of control $\{u_n\}$,

$$V_0(X_0) \leq g(X_N) + \sum_{n=0}^{N-1} f_n(X_n, u_n).$$

Taking infimum on the RHS over all controls, we have $V_0(X_0) \leq v(X_0)$. Again, similar to Example 1, if there exists a sequence $\{u_n^* \doteq u_n^*(x) \in U_n(x)\}$ attaining the infimum of the RHS of the DPE, then

$$V_0(X_0) = g(X_N^*) + \sum_{n=0}^{N-1} f_n(X_n^*, u_n^*).$$

In particular, $V_0(X_0) = v(X_0)$ and $\{u_n^*\}$ is optimal.

It remains to show that $V_0(X_0) \geq v(X_0)$ even if we do not know the existence of $\{u_n^*\}$ a prior. Let $\varepsilon > 0$ be an arbitrary small positive real number. By definition, there exists a $u_0^\varepsilon \in U_0(X_0)$ such that

$$V_0(X_0) \geq f_0(X_0, u_0^\varepsilon) + V_1\left(h_0(X_0, u_0^\varepsilon)\right) - \varepsilon \doteq f_0(X_0, u_0^\varepsilon) + V_1\left(X_1^\varepsilon\right) - \varepsilon.$$

Again by definition, there exists a $u_1^\varepsilon \in U_1(X_1^\varepsilon)$ such that

$$V_1(X_1^\varepsilon) \geq f_1(X_1^\varepsilon, u_1^\varepsilon) + V_2\left(h_1(X_1^\varepsilon, u_1^\varepsilon)\right) - \varepsilon \doteq f_1(X_1^\varepsilon, u_1^\varepsilon) + V_2\left(X_2^\varepsilon\right) - \varepsilon.$$

Repeating this procedure, we arrive at (with $X_0^\varepsilon \doteq X_0$)

$$V_0(X_0) \geq g(X_N^\varepsilon) + \sum_{n=0}^{N-1} f_n(X_n^\varepsilon, u_n^\varepsilon) - N\varepsilon \geq v(X_0) - N\varepsilon.$$

Letting $\varepsilon \to 0$, we complete the proof. ■

The next example is a continuous-time deterministic optimal control problem. The idea of the DP is similar, but the resulting DPE is a *partial differential equation* (PDE).

Example 3 (Continuous-time utility maximization). Consider an agent with initial wealth X_0. The money is put in a bank which will offer an (compound) interest rate of $r > 0$ per unit of time. At any time $t \in [0, T]$, the agent withdraws part of the wealth for consumption. Let the total consumption up until time t is C_t and assume that C_t has the form

$$C_t = \int_0^t c_s \, ds = \int_0^t u_s X_s \, ds, \quad \forall\, t \in [0, T].$$

Naturally, we should put a constraint $u_t \geq 0$ for every t. The corresponding wealth process has the dynamics

$$dX_t = rX_t \, dt - u_t X_t \, dt, \qquad \forall\, t \in [0, T].$$

The optimization for the agent is to come up with a consumption process $\{u_t\}$ so as to maximize the associated total utility

$$\sup_{\{u_t\}} \left[\int_0^T f(u_t X_t) \, dt + g(X_T) \right].$$

How should the agent proceed to find this optimal consumption process?

The solution is again divided into two steps. However, as we will see, the second step (the verification) needs some extra regularity conditions. The reason is that the existence of a solution to a differential equation, unlike a difference equation (from discrete-time model), is not automatic.

Step 1: This step is to intuitively derive the DPE. The basic idea of DP is the same as that of the discrete-time model. However, one now has the freedom to choose time steps. Recall the interpretation of V_n as the maximal utility from time n. Analogously, assume for now that $V(t; x)$ is the maximal utility from time t. Let δ be a small positive number. What should the agent do in a time interval of length δ, say $[t, t+\delta]$? Apparently, it is optimal to choose a control policy so as to maximize

$$\int_t^{t+\delta} f(u_s X_s) \, ds + V(t+\delta; X_{t+\delta}).$$

The maximal value will equal $V(t; X_t)$. In other words, one expects

$$V(t; X_t) = \sup_{\{u_s : s \in [t, t+\delta]\}} \left[\int_t^{t+\delta} f(u_s X_s) \, ds + V(t+\delta; X_{t+\delta}) \right].$$

Now let $u \geq 0$ be an arbitrary constant, and consider a specific control by $u_s \equiv u$ for $s \in [t, t+\delta]$. It follows that

$$V(t; X_t) \geq \int_t^{t+\delta} f(u X_s) \, ds + V(t+\delta; X_{t+\delta}),$$

or

$$0 \geq \frac{1}{\delta} \int_t^{t+\delta} f(uX_s)\, ds + \frac{V(t+\delta; X_{t+\delta}) - V(t;x)}{\delta}.$$

Letting $\delta \to 0$ and assuming that $V(t;x)$ is continuously differentiable with respect to t and x, we have

$$0 \geq \left[f(ux) + \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}(rx - ux) \right] \Big|_{x=X_t}. \tag{9}$$

However, suppose the optimal control $\{u_s^* : s \in [t, t+\delta]\}$ exists, then

$$V(t; X_t) = \int_t^{t+\delta} f(u_s^* X_s^*)\, ds + V(t+\delta; X_{t+\delta}^*).$$

Subtracting both sides by $V(t, X_t)$ and dividing by δ, one would expect (under some minor continuity conditions) that

$$0 = \left[f(u_t^* x) + \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}(rx - u_t^* x) \right] \Big|_{x=X_t} \tag{10}$$

as $\delta \to 0$. Combining equations (9) and (10), and recalling that the argument holds no matter what value X_t is, one arrive at a non-linear PDE

$$\begin{aligned} \frac{\partial V}{\partial t} + \sup_{u \geq 0} \left[\frac{\partial V}{\partial x}(rx - ux) + f(ux) \right] &= 0 && (11) \\ V(T; x) &= g(x). && (12) \end{aligned}$$

This equation is called the *Hamilton-Jacobi-Bellman* (HJB) equation, with the terminal condition (12) being quite self-explanatory.

One would expect that the solution to the HJB equation, say (abusing the notation a bit) V, should yield the value function; more precisely,

$$v(x) = V(0; x)$$

and the optimization u^* in this equation (may depend on x and t) should lead to an optimal control.

Step 2: The HJB equation (11)-(12) is nonlinear, and there may not exist a (classical) solution. But when there does exist a (classical) solution, usually the verification will work. We should assume the following.

Condition **1** *1. The HJB equation (11)-(12) admits a solution V.*

2. *Suppose the optimizer in the HJB equation is $u^*(t;x)$. Then the equation*
$$dX_t^* = rX_t^*\,dt - u^*(t;X_t)X_t^*\,dt, \quad X_0^* = X_0$$
has a solution.

We will show that, under Condition 1, $v(X_0) = V(0;X_0)$, and $\{u_t^* \doteq u^*(t;X_t^*)\}$ is an optimal control.

Let $\{u_t\}$ be an arbitrary consumption policy. Then

$$\begin{aligned} V(T;X_T) - V(0;X_0) &= \int_0^T \left[\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}(rx - ux)\Big|_{x=X_t,u=u_t} \right] dt \\ &\le \int_0^T \left[-f(u_tX_t)\right] dt. \end{aligned}$$

Thus

$$V(0;X_0) \ge V(T;X_T) + \int_0^T f(u_tX_t)\,dt = g(X_T) + \int_0^T f(u_tX_t)\,dt.$$

Taking supremum over all consumption policy $\{u_t\}$, we arrive at

$$V(0;X_0) \ge v(X_0).$$

Now plugging in $\{u_t^*\}$ and $\{X_t^*\}$, all inequalities becomes equality, and we have
$$V(0;X_0) = g(X_T^*) + \int_0^T f(u_t^*X_t^*)\,dt.$$
It follows readily that $V(0;X_0) = v(X_0)$ and $\{u_t^*\}$ is an optimal consumption policy. ■.

Remark **3** In general, if we define $V(t;x)$ as the maximal utility from time t given $X_t = x$, then V is not continuously differentiable. Therefore, it cannot satisfy the HJB equation in the classical sense. However, under very mild conditions, it can be shown that V satisfies the HJB equation in a *viscosity* sense; see [1].

Example 4 Obtaining a closed-form solution to the non-linear HJB equation is usually impossible. But in some special cases, it is possible. In this example, we will work out the details for the case where f, g are power utilities; i.e., $f(x) = g(x) = x^\alpha$ for some $\alpha \in (0,1)$.

The corresponding HJB equation becomes

$$\begin{aligned} \frac{\partial V}{\partial t} + \sup_{u \geq 0} \left[\frac{\partial V}{\partial x}(rx - ux) + u^\alpha x^\alpha \right] &= 0 \\ V(T;x) &= x^\alpha. \end{aligned}$$

Guess that $V(t;x)$ has form $V(t;x) = f(t)x^\alpha$. It is easy to check that the above PDE reduces to an ordinary differential equation (ODE)

$$\begin{aligned} f'(t) + \alpha r f(t) + (1-\alpha)\,[f(t)]^{\frac{\alpha}{\alpha-1}} &= 0 \\ f(T) &= 1, \end{aligned}$$

and the optimizing $u^*(t;x)$ is independent of x; i.e.,

$$u^*(t;x) = [f(t)]^{\frac{1}{\alpha-1}} \doteq u^*(t).$$

The solution to the ODE takes form

$$f(t) = \left[\left(\frac{1-\alpha}{\alpha r} + 1 \right) e^{\frac{\alpha r}{1-\alpha}(T-t)} - \frac{1-\alpha}{\alpha r} \right]^{1-\alpha}$$

Therefore, the HJB equation has a classical solution of form $V(t;x) = f(t)x^\alpha$, and clearly the equation

$$dX_t^* = rX_t^*\,dt - u^*(t)X_t^*\,dt = (r - u^*(t))X_t^*\,dt$$

has a solution. Thus Condition 1 is satisfies, and

$$v(X_0) = V(0;X_0) = f(0)X_0^\alpha.$$

We conclude that the maximal utility is of power type, if both f and g are of power type, and the optimal consumption policy is independent of the current wealth. ■

References

[1] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations.* Birkhäuser, 1997.

[2] D. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete Time Case.* Academic Press, San Diego, California, 1978.

[3] D.P. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models.* Prentice Hall, Englewood Cliffs, NJ, 1987.

[4] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA, 1995.

[5] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA, 1995.

[6] C. Derman. *Finite State Markovian Decision Processes*. Academic press, New York, 1970.

[7] W. H. Fleming and R. Rishel. *Deterministic and Stochastic Optimal Control*. Springer, Berlin, 1975.

[8] W. H. Fleming and H. M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, New York, 1992.

[9] A.N. Shiryaev. *Optimal Stopping Rules*. Springer, Berlin, 1978.

[10] N.L. Stokey, R. Lucas, and E. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, MA, 1989.