3 Finite-horizon dynamic optimization

We show how to optimize a controllable dynamical system over finitely many time steps. Fix a time horizon $n \in \mathbb{Z}^+$ and assume that

$$c(n, x, a) = C(x) \quad \text{and} \quad c(k, x, a) = 0, \quad k \geqslant n+1, \quad x \in S, \quad a \in A.$$

Thus the total cost function is given by

$$V^u(k, x) = \sum_{j=k}^{n-1} c(j, x_j, u_j) + C(x_n), \quad 0 \leqslant k \leqslant n,$$

in the deterministic case, and in the stochastic case by

$$V^u(k, x) = \mathbb{E}^u_{(k,x)} \left(\sum_{j=k}^{n-1} c(j, X_j, U_j) + C(X_n) \right), \quad 0 \leqslant k \leqslant n.$$

Note that $V(k, x) = 0$ for all $k \geqslant n+1$. Hence, the optimality equation can be written in the form

$$\begin{aligned} V(n, x) &= C(x), && x \in S, \\ V(k, x) &= \inf_a \{c(k, x, a) + V(k+1, f(k, x, a))\}, \quad & 0 \leqslant k \leqslant n-1, \quad & x \in S, \end{aligned}$$

in the deterministic case, and in the stochastic case by[12]

$$\begin{aligned} V(n, x) &= C(x), && x \in S, \\ V(k, x) &= \inf_a (c + PV)(k, x, a), \quad & 0 \leqslant k \leqslant n-1, \quad & x \in S. \end{aligned}$$

Both these equations have a unique solution, which moreover may be computed by a straightforward[13] backwards recursion from time n. Once we have computed V, an optimal control can be identified whenever we can find a minimizing action in the optimality equations for $0 \leqslant k \leqslant n-1$. The following easy result verifies this for the deterministic case.

[12] It is often convenient to write the equation in terms of the *time to go* $m = n - k$. Assume that P is time-homogeneous and set $V_m(x) = V(k, x)$ and $c_m(x, a) = c(k, x, a)$, then the optimality equations become $V_0(x) = C(x)$ and

$$V_{m+1}(x) = \inf_a (c_m + PV_m)(x, a), \quad 0 \leqslant m \leqslant n-1, \quad x \in S.$$

In particular, in the case where both P and c are time-homogeneous, if we define

$$V^u_n(x) = \mathbb{E}^u_x \sum_{j=0}^{n-1} c(X_j, U_j), \quad V_n(x) = \inf_u V^u_n(x),$$

then the functions V_n are given by $V_0(x) = 0$ and, for $n \geqslant 0$,

$$V_{n+1}(x) = \inf_a (c + PV_n)(x, a), \quad x \in S.$$

[13] Although straightforward in concept, the size of the state space may make this a demanding procedure in practice. It is worth remembering, as a possible alternative, the following *interchange argument*, when

Proposition 3.1. *Suppose we can find a control u, with controlled sequence $(x_0, \ldots, x_n)$ such that*

$$V(k, x_k) = c(k, x_k, u_k) + V(k+1, f(k, x_k, u_k)), \quad 0 \leqslant k \leqslant n-1.$$

Then u is optimal for $(0, x_0)$.

Proof. Fix a such a control u, and set

$$m_k = \sum_{j=0}^{k-1} c(j, x_j, u_j) + V(k, x_k), \quad 0 \leqslant k \leqslant n.$$

Then, for $0 \leqslant k \leqslant n-1$, since $x_{k+1} = f(k, x_k, u_k)$, we have

$$m_{k+1} - m_k = c(k, x_k, u_k) + V(k+1, x_{k+1}) - V(k, x_k) = 0.$$

Hence

$$V(0, x_0) = m_0 = m_n = \sum_{j=0}^{n-1} c(j, x_j, u_j) + C(x_n).$$

□

Example (Managing spending and saving). An investor holds a capital sum in a building society, which pays a fixed rate of interest $\theta \times 100\%$ on the sum held at each time $k = 0, 1, \ldots, n-1$. The investor can choose to reinvest a proportion a of the interest paid, which then itself attracts interest. No amounts invested can be withdrawn. How should the investor act to maximize total consumption by time $n-1$?

Take as state the present income $x \in \mathbb{R}^+$ and as action the proportion $a \in [0, 1]$ which is reinvested. The income next time is then

$$f(x, a) = x + \theta a x$$

and the reward this time is $r(x, a) = (1-a)x$. The optimality equation is given by

$$V(k, x) = \max_{0 \leqslant a \leqslant 1} \{(1-a)x + V(k+1, (1+\theta a)x)\}, \quad 0 \leqslant k \leqslant n-1,$$

---

seeking to optimize the order in which one performs a sequence of n tasks. Label the tasks $\{1, \ldots, n\}$ and write $c(\sigma)$ for the cost of performing the tasks in the order $\sigma = (\sigma_1, \ldots, \sigma_n)$. We examine the effect on $c(\sigma)$ of interchanging the order of two of the tasks. Suppose we can find a function f on $\{1, \ldots, n\}$ such that, for all σ and all $0 \leqslant i \leqslant n-1$,

$$c(\sigma') < c(\sigma) \quad \text{whenever} \quad f(\sigma_i) > f(\sigma_{i+1}),$$

where σ' is obtained from σ by interchanging the order of tasks σ_i and σ_{i+1}. Then the condition $f(\sigma_1) \leqslant \ldots \leqslant f(\sigma_n)$ is necessary for optimality of σ. This may be enough to reduce the number of possible optimal orders to 1. In any case, if we have also, for all σ and all $0 \leqslant i \leqslant n-1$,

$$c(\sigma') = c(\sigma) \quad \text{whenever} \quad f(\sigma_{i+1}) = f(\sigma_i),$$

then our optimality condition is also sufficient.

with $V(n,x) = 0$. Working back from time n, we see that $V(k,x) = c_{n-k}x$ for some constants $c_0, \dots, c_n$, given by $c_0 = 0$ and

$$c_{m+1} = \max\{c_m + 1, (1+\theta)c_m\}, \quad 0 \leqslant m \leqslant n-1.$$

Hence

$$c_m = \begin{cases} m, & m \leqslant m^*, \\ m^*(1+\theta)^{m-m^*}, & m > m^*, \end{cases}$$

where $m^* = \lceil 1/\theta \rceil$. By Proposition 3.1, the optimal control is to reinvest everything before time $k^* = n - m^*$ and to consume everything from then on.

The optimality of a control in the stochastic case can verified using the following result.

Proposition 3.2. *Suppose we can find a Markov control u such that*

$$V(k,x) = (c + PV)(k,x,u_k(x)), \quad 0 \leqslant k \leqslant n-1, \quad x \in S.$$

Then u is optimal for all (k,x).

Proof. Fix such a Markov control u and write $(X_0, \dots, X_n)$ for the associated Markov chain starting from $(0,x)$. Define

$$M_k = \sum_{j=0}^{k-1} c(j, X_j, U_j) + V(k, X_k), \quad 0 \leqslant k \leqslant n.$$

Then, for $0 \leqslant k \leqslant n-1$,

$$M_{k+1} - M_k = c(k, X_k, U_k) + V(k+1, X_{k+1}) - V(k, X_k),$$

so, for all $y \in S$,

$$\mathbb{E}^u(M_{k+1} - M_k | X_k = y) = (c + PV)(k, y, u_k(y)) - V(k,y) = 0.$$

Hence

$$V(0,x) = \mathbb{E}_x^u(M_0) = \mathbb{E}_x^u(M_n) = \mathbb{E}_x^u\left(\sum_{j=0}^{n-1} c(j, X_j, U_j) + C(X_n)\right).$$

The same argument works for all starting times k. □

Example (Exercising a stock option). You hold an option to buy a share at a fixed price p, which can be exercised at any time $k = 0, 1, \dots, n-1$. The share price satisfies $Y_{k+1} = Y_k + \varepsilon_{k+1}$, where $(\varepsilon_k)_{k \geqslant 1}$ is a sequence of independent identically distributed random variables[14], with $\mathbb{E}(|\varepsilon|) < \infty$. How should you act to maximise your expected return?

Take as state the share price $x \in \mathbb{R}$, until we exercise the option, when we move to a terminal state ∂. Take as action space the set $\{0, 1\}$, where $a = 1$ corresponds to exercising the option. The problem specifies a realised stochastic controllable dynamical system. We

[14] Thus we allow, unrealistically, the possibility that the price could be negative. This model might perhaps be used over a small time interval, with Y_0 large.

are working outside the countable framework here, but in the realised case, where PV is given straightforwardly by 1. The rewards and dynamics before termination are given by

$$r(x,a) = a(x-p), \quad G(x,a,\varepsilon) = \begin{cases} x+\varepsilon, & \text{if } a=0, \\ \partial, & \text{if } a=1, \end{cases}.$$

Hence the optimality equation is given by

$$V(k,x) = \max\{x-p, \mathbb{E}(V(k+1, x+\varepsilon))\}, \quad k = 0,1,\dots,n-1,$$

with $V(n,x) = 0$. Note that $V(n-1,x) = (x-p)^+$. By backwards induction, we can show that $V(k,.)$ is convex for all k, and increases as k decreases. Set $p_k = \inf\{x \geqslant 0 : V(k,x) = x-p\}$. Then p_k increases as k decreases and the optimal control is to exercise the option as soon as $Y_k = p_k$.