5. Dynamic optimization for discounted costs

We are given

$$P : S \times A \longrightarrow \text{Prob}(S)$$

and a bounded function

$$c : S \times A \longrightarrow [-C, C]$$

Future costs are discounted by $\beta \in (0,1)$ per time step.
In the framework of §2 our cost function is

$$(n, x, a) \longmapsto \beta^n c(x, a)$$

when starting at time 0.

Discounting is normal in financial models, reflecting the fact that money can be invested to earn interest

Define, for a control u, the <u>expected discounted cost function</u>

$$V^u(x) = \mathbb{E}_0^x \sum_{n=0}^{\infty} \beta^n c(X_n, U_n)$$

and the <u>infimal discounted cost function</u>

$$V(x) = \inf_u V^u(x).$$

We have

$$\sum_{n=0}^{\infty} |\beta^n c(X_n, U_n)| \leq C \sum_{n=0}^{\infty} \beta^n = \frac{C}{1-\beta} < \infty$$

This allows us to follow the arguments in §2 without assuming $c \geq 0$ or $c \leq 0$.

Value iteration

Consider

$$V_n^u(x) = \mathbb{E}_x^u \sum_{k=0}^{n-1} \beta^k c(X_k, U_k), \qquad V_n(x) = \inf_u V_n^u(x).$$

We have

$$|V_n^u(x) - V^u(x)| \leq C \sum_{k=n}^{\infty} \beta^k = \frac{C\beta^n}{1-\beta} \longrightarrow 0 \quad \text{as } n \to \infty$$

so

$$V_n^u(x) - \frac{C\beta^n}{1-\beta} \leq V^u(x) \leq V_n^u(x) + \frac{C\beta^n}{1-\beta}$$

and, on taking the infimum over u

$$V_n(x) - \frac{C\beta^n}{1-\beta} \leq V(x) \leq V_n(x) + \frac{C\beta^n}{1-\beta},$$

so

$$V_n \longrightarrow V \quad \text{uniformly on } S.$$

The finite-horizon cost functions V_n are determined iteratively by the optimality equations

$$V_0(x) = 0, \qquad V_{n+1}(x) = \inf_a (c + \beta P V_n)(x, a), \quad x \in S.$$

So, as in the case of non-negative rewards, we can compute V as the limit of this iteration.

(In §2, fix m and take

$$c(n,x,a) = \beta^n c(x,a) 1_{n \le m}.$$

Then, using time-homogeneity,

$$V(0,x) = V_{m+1}(x), \quad V(1,x) = \beta V_m(x)$$

so, by Proposition 2.1,

$$V_{m+1}(x) = V(0,x) = \inf_a (c + PV)(0,x,a)$$

$$= \inf_a \{ c(x,a) + \sum_y P(x,a)_y \, \beta V_m(y) \}$$

$$= \inf_a (c + \beta P V_m)(x,a).$$

Similarly, taking $m = \infty$, $V(x) = \inf_a (c + \beta P V)(x,a)$)

Proposition 5.1

The infimal discounted cost function is the unique bounded solution of the dynamic optimality equation

$$V(x) = \inf_a (c + \beta P V)(x, a), \quad x \in S.$$

Moreover, any map $u : S \to A$ such that

$$V(x) = (c + \beta P V)(x, u(x)), \quad x \in S$$

defines a stationary Markov control, which is optimal in the class of all controls, for all starting states x.

Proof We have seen already that V is bounded and satisfies the optimality equation. Suppose F is another bounded solution of the optimality equation and u is any control. Define, for a (P,u)-controlled process $(X_n)_{n\geq 0}$

$$M_n = \sum_{k=0}^{n-1} \beta^k c(X_k, U_k) + \beta^n F(X_n), \quad n \geq 0.$$

Then

$$M_{n+1} - M_n = \beta^n c(X_n, U_n) + \beta^{n+1} F(X_{n+1}) - \beta^n F(X_n),$$

so, for all $y \in S$ and $a \in A$,

$$E_x^u(M_{n+1} - M_n \mid X_n = y, U_n = a) = \beta^n c(y,a) + \beta^{n+1} PF(y,a) - \beta^n F(y) \geq 0$$

Hence

$$F(x) = M_0 \leq \mathbb{E}_x^u(M_n) = V_n^u(x) + \beta^n \mathbb{E}_x^u F(X_n).$$

On letting $n \to \infty$, recalling that F is bounded, we obtain $F \leq V^u$. Since u was arbitrary, this implies $F \leq V$.

*

If u can be found as in the statement, then

$$\mathbb{E}_x^u(M_{n+1} - M_n \mid X_n = y) = 0, \quad y \in S,$$

so

$$F(x) = M_0 = \mathbb{E}_x^u(M_n) = V_n^u(x) + \beta^n \mathbb{E}_x^u F(X_n).$$

Letting $n \to \infty$, this gives $F = V^u$. Since $F \leq V \leq V^u$, we deduce that $F = V$ and u is optimal.

In general such a control u may not exist. But, given $\varepsilon > 0$, there does exist $\tilde{u} : S \to A$ such that

$$(c + \beta P F)(x, \tilde{u}(x)) \leq F(x) + \varepsilon \quad , \quad x \in S,$$

which we can write as

$$F(x) = (\tilde{c} + \beta P F)(x, \tilde{u}(x)), \quad x \in S,$$

for a new cost function $\tilde{c} \geq c - \varepsilon$. Arguing as above, with $\tilde{c}, \tilde{u}$ in place of c, u, we get

$$F(x) = \mathbb{E}_x^{\tilde{u}} \sum_{k=0}^{\infty} \beta^k \tilde{c}(X_k, \tilde{u}(X_k)) \geq V^{\tilde{u}}(x) - \frac{\varepsilon}{1-\beta} \geq V(x) - \frac{\varepsilon}{1-\beta},$$

But ε was arbitrary, so $F \geq V$ and so $F = V$.

□

Example – When is the right time to sell?

A speculator owns a rare collection of tulip bulbs.
Each day, he has one opportunity to sell the collection.
The prices offered each day are independent, with density function $p(x)$.
Also, each day, there is a probability $1-\beta < \frac{1}{2}$ that the entire collection will succumb to disease, rendering it worthless.
How should the speculator maximize his expected return?
Obtain the optimal control explicitly when $p(x) = \frac{2}{x^3}, x \geq 1$

6. Dynamic optimization for non-negative costs

(also called negative programming)

We are given

$$P : S \times A \longrightarrow \mathrm{Prob}(S), \qquad c : S \times A \longrightarrow \mathbb{R}^+.$$

Let $u : S^* \longrightarrow A$ be a control, and $(X_n)_{n \geq 0}$ a (P, u)-controlled process starting from x at time 0.

Define

$$V^u(x) = \mathbb{E}_x^u \sum_{n=0}^{\infty} c(X_n, U_n) \qquad \text{expected total cost function}$$

$$V(x) = \inf_u V^u(x) \qquad \text{infimal cost function}$$

By monotone convergence

$$V^u(x) = \lim_{n \to \infty} V_n^u(x) \quad \text{for} \quad V_n^u(x) = \mathbb{E}_x^u \sum_{k=0}^{n-1} c(X_k, U_k)$$

Proposition 6.1

Assume that A is finite. Then the infimal cost function is the minimal non-negative solution of the optimality equation

$$V(x) = \min_a (c + PV)(x,a), \quad x \in S.$$

Moreover, any map $u : S \to A$ such that

$$V(x) = (c + PV)(x, u(x)), \quad x \in S$$

defines a stationary Markov control which is optimal in the class of all controls, for every starting state x.

(Recall $PV(x,a) = \sum_y P(x,a)_y V(y)$.)

Proof Time homogeneity and Proposition 2.1 combine to show that V satisfies the optimality equation.

Suppose F is another non-negative solution.

Since A is finite, there exists $\tilde{u} : S \to A$ such that

$$F(x) = (C + PF)(x, \tilde{u}(x)), \quad x \in S.$$

By an argument used in Proposition 5.1 (taking $\beta = 1$)

$$F(x) = V_n^{\tilde{u}}(x) + \mathbb{E}_x^{\tilde{u}} F(X_n) \geq V_n^{\tilde{u}}(x)$$

On letting $n \to \infty$, we obtain $F \geq V^{\tilde{u}} \geq V$, so V is the minimal non-negative solution.

For u as in the statement, we can take $F = V$, $\hat{u} = u$ in this argument to see that $V \geq V^u$, so u is optimal. □

Value iteration

Set

$$V_n(x) = \inf_u V_n^u(x), \quad V_\infty(x) = \lim_{n\to\infty} V_n(x), \quad x \in S$$

Since $V_n^u \le V^u$ for all n and u, we can take the infimum over controls to obtain $V_n \le V$ and hence $V_\infty \le V$.

On the other hand

$$V_{n+1}(x) = \min_a (C + PV_n)(x,a), \quad x \in S$$

so, letting $n \to \infty$, by monotone convergence,

$$V_\infty(x) = \min_a (C + PV_\infty)(x,a), \quad x \in S$$

so, by Proposition 6.1, $V \le V_\infty$ and hence $V_\infty = V$.

Thus we can again compute V as a limit of an iteration.

Policy improvement

For any stationary Markov control $u : S \to A$ we have

$$V^u_{n+1}(x) = (C + PV_n)(x, u(x)), \quad V^u(x) = (C + PV)(x, u(x)), \quad x \in S.$$

(These equations come from Markov chain arguments – conditioning on the first step, or, if you like they are special cases of the optimality equation, when $|A| = 1$.)

If V^u does not satisfy the optimality equation, then there exists $\tilde{u} : S \to A$ such that

$$V^u(x) \geq (C + PV^u)(x, \tilde{u}(x)), \quad x \in S$$

with strict inequality at some $x_0 \in S$.

Then, certainly $V^u \geq V_0^{\tilde{u}} = 0$.

Suppose inductively that $V^u \geq V_n^{\tilde{u}}$. Then

$$V^u(x) \geq (c + PV^u)(x, \tilde{u}(x)) \geq (c + PV_n^{\tilde{u}})(x, \tilde{u}(x)) = V_{n+1}^{\tilde{u}}(x)$$

so the induction proceeds.

Letting $n \to \infty$, we get $V^u \geq V^{\tilde{u}}$, with strict inequality at x_0, so we have found an improved control $\tilde{u}$.

Summary

Non-negative rewards

- V is the minimal non-negative solution of OE
- V^u satisfies OE $\Rightarrow$ u optimal
- minimizing actions in OE may fail to give optimal control

Discounted bounded costs or rewards

- V is the unique bounded solution of OE
- minimizing actions in OE give optimal (Markov) control

Non-negative costs, A finite

- V is the minimal non-negative solution of OE
- minimizing actions in OE give optimal (Markov) control

7. Optimal stopping

We are given a Markov chain $(X_n)_{n\geq 0}$ in S, with transition matrix $P = (p_{xy} : x, y \in S)$ say, and two bounded functions

$$c : S \to \mathbb{R} \quad \underline{\text{continuation cost}}, \qquad f : S \to \mathbb{R} \quad \underline{\text{stopping cost}}.$$

A random variable T, with values in $\mathbb{Z}^+ \cup \{\infty\}$, is a $\underline{\text{stopping time}}$ if, for all $n \geq 0$, there exists $B_n \subseteq S^{n+1}$ such that

$$\{T = n\} = \{(X_0, \ldots, X_n) \in B_n\}.$$

Thus, the stopping times are all the rule for stopping based on the history of X up to the present time.

Define the expected total cost function

$$V^T(x) = \mathbb{E}_x\left(\sum_{k=0}^{T-1} c(X_k) + f(X_T)\mathbb{1}_{\{T<\infty\}}\right), \quad x \in S,$$

and define for $n \in \mathbb{Z}^+$ and $x \in S$,

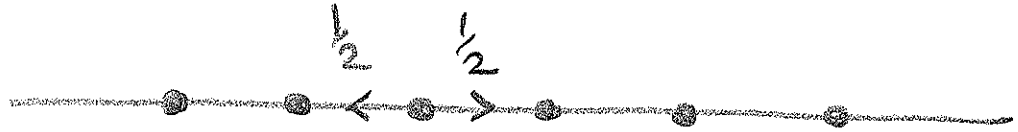
$$V_n(x) = \inf_{T \le n} V^T(x), \quad V_*(x) = \inf_{T<\infty} V^T(x), \quad V(x) = \inf_T V^T(x)$$

where the infima are taken over all stopping times, first with the restriction $T \le n$, then just $T < \infty$, then unrestricted.

In the second and third cases we shall assume $c, f \ge 0$.

Note that $V_n \le V_{n+1} \le V_* \le V$.

Example



Take $c \equiv 0$, $f(x) = 1 + e^{-x}$. Certainly $V_0 = f$.

Suppose inductively that $V_n = f$.

By conditioning on the first step,

$$V_{n+1}(x) = \min\{f(x),\ \tfrac{1}{2}V_n(x+1) + \tfrac{1}{2}V_n(x-1)\} = f(x)$$

since f is convex, and the induction proceeds.

We know that $(X_n)_{n\geq 0}$ is recurrent, so

$$T_N = \inf\{n \geq 0 : X_n = N\} < \infty \quad \text{(with probability 1)}.$$

Hence

$$V_*(x) \leq V^{T_N}(x) = 1 + e^{-N} \quad \text{for all } N$$

Certainly $V_* \geq 1$ so $V_* = 1$.

On the other hand the choice $T \equiv \infty$ shows $V(x) = 0$.

In this example

$$\inf_n V_n(x) > V_*(x) > V(x) \quad \text{for all } x.$$

Proposition 7.1 (One step look ahead rule)

Suppose that $(X_n)_{n\geq 0}$ cannot escape from the set

$$S_0 = \{x \in S : f(x) \leq (c + Pf)(x)\}.$$

Then, for all $n \geq 0$, the following stopping time is optimal for the n-horizon problem

$$T_n = \inf\{k \geq 0 : X_k \in S_0\} \wedge n.$$

Proof The claim holds for $n=0$. Let us suppose it holds for n. Then $V_n = f$ on S_0, so $PV_n = Pf$ on S_0 as we cannot escape. So, for $x \in S_0$

$$V_{n+1}(x) = \min\{f(x), (c + PV_n)(x)\} = f(x)$$

and we should stop immediately.

On the other hand, for $x \notin S_0$,

$$V_{n+1}(x) = \min\{f(x), (c + PV_n)(x)\} = (c + PV_n)(x)$$

$$\left(\leq (c + Pf)(x) < f(x) \right)$$

and it is better to wait, at least one step.

By the Markov property and our inductive hypothesis, in the n steps remaining it is optimal to wait until we hit S_0, or time runs out.

Hence the claim holds for all n.

Example - optimal parking

You intend to park on the Backs, and wish to minimize the expected distance you will have to walk to Garrett Hostel Lane.

Suppose each parking space is free, independently, with probability $p \in (0, 1)$.

Assume that a queue of cars behind you take up immediately any space you pass by, and that no new spaces are vacated.

Where should you park?

We have been using the equation

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

This can be derived carefully by translating the optimal stopping problem into a dynamic optimization problem.

Take as state-space $S \cup \{\partial\}$, action space $A = \{0, 1\}$.

Action 1 will correspond to stopping.

On stopping we go to ∂ and stay there. Define

$$P(x,0)_y = P_{xy}, \quad P(x,1)_\partial = \delta_{y\partial}$$

$$c(x,a) = \begin{cases} c(x), & a = 0 \\ f(x), & a = 1 \end{cases} \qquad c(\partial, a) = 0$$

Given a stopping time T, we can define a control $u : S_0^* \to A$

$$u_n(x_0, \ldots, x_n) = \begin{cases} 1 & \text{if } (x_0, \ldots, x_n) \in B_n \\ 0 & \text{otherwise.} \end{cases}$$

We get all controls $u : S_0^* \to A$ in this way.

The (P, u)-controlled process $(\tilde{X}_n)_{n \geq 0}$ is given by

$$\tilde{X}_n = \begin{cases} X_n, & n \leq T, \\ \partial, & n > T. \end{cases}$$

Hence V_n is exactly the n-horizon infimal cost function with final cost f. From §3, we have then $V_0 = f$ and

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

8. Dynamic optimization for long-run average costs

We are given

$$P : S \times A \longrightarrow \mathrm{Prob}(S), \qquad c : S \times A \longrightarrow [-C, C].$$

Given a control $u : S^* \longrightarrow A$, let $(X_n)_{n \geq 0}$ be a (P, u)-controlled process, starting from x, set $U_n = u_n(X_0, \ldots, X_n)$.

Define

$$V_n^u(x) = \mathbb{E}_x^u \sum_{k=0}^{n-1} c(X_k, U_k).$$

Say that u is optimal, starting from x, if $\lambda = \lim_{n \to \infty} V_n^u(x)/n$ exists and if, for all other controls $\tilde{u}$, $\lambda \leq \liminf_{n \to \infty} V_n^{\tilde{u}}(x)/n$.

Then λ is the minimal long-run average cost, starting from x

Proposition 8.1

Suppose there exists a constant λ and a bounded function Θ on S such that

$$\lambda + \Theta(x) \leq (c + P\Theta)(x,a), \quad x \in S, \ a \in A.$$

Then, for all controls u and all $x \in S$

$$\liminf_{n \to \infty} V_n^u(x)/n \geq \lambda .$$

(Recall that $P\Theta(x,a) = \sum_y P(x,a)_y \Theta(y)$)

Proof Fix u and consider

$$M_n = \Theta(X_n) + \sum_{k=0}^{n-1} c(X_k, U_k) - n\lambda .$$

Then

$$M_{n+1} - M_n = \Theta(X_{n+1}) - \Theta(X_n) + c(X_n, U_n) - \lambda$$

so, for all $y \in S$ and $a \in A$,

$$E_x^u(M_{n+1} - M_n \mid X_n = y, U_n = a) = P\Theta(y,a) - \Theta(y) + c(y,a) - \lambda \geq 0.$$

Hence

$$\Theta(x) = M_0 \leq E_x^u(M_n) = E_x^u(\Theta(X_n)) - n\lambda + V_n^u(x)$$

and so

$$\frac{V_n^u(x)}{n} \geq \lambda + \frac{\Theta(x)}{n} - \frac{E_n^u(\Theta(X_n))}{n} \to \lambda \quad \text{as } n \to \infty .$$

□

Proposition 8.2

Suppose that there exists a constant λ, a bounded function Θ on S, and a map $u : S \to A$ such that

$$\lambda + \Theta(x) \geq (C + P\Theta)(x, u(x)), \quad x \in S.$$

Then, for all $x \in S$,

$$\limsup_{n \to \infty} \frac{V_n^u(x)}{n} \leq \lambda.$$

This can be proved along the same lines as Proposition 8.1.

Proposition 8.3

Suppose that there exists a constant λ and a bounded function Θ on S satisfying the dynamic optimality equation

$$\lambda + \Theta(x) = \inf_a (c + P\Theta)(x,a), \quad x \in S,$$

and suppose the infimum is achieved at $u(x)$ for each x. Then u defines a stationary Markov control, which is optimal in the class of all controls, and λ is the minimal long-run average cost, simultaneously for all starting points x.

Proof By Proposition 8.1, for all controls $\tilde{u}$ and all $x \in S$

$$\liminf_{n \to \infty} V_n^{\tilde{u}}(x)/n \geq \lambda.$$

On the other hand, for the control u, we can replace all inequalities by equalities in the proof of Proposition 8.1, to obtain

$$\lim_{n \to \infty} V_n^{u}(x)/n = \lambda .$$

□

What is Θ ?

Note that $\Theta + 19$ works as well as Θ.

The function Θ measures the relative cost of starting at each state.

Example – consultant's job selection

Each day, a consultant is either free or is occupied with some job, which may be of n different types $x = 1, \ldots, n$. Whenever he is free, he is given the opportunity to take on a new job starting the next day.

A job of type x is offered with probability π_x and the type of job offered on each day is independent.

On any day when he work on a job of type x, he completes it with probability p_x, independently for each day, and on its completion he is paid R_x.

Which jobs should he accept?

Value iteration

Define V_n, $n \geq 0$, by the finite-horizon optimality equations

$$V_0(x) = 0, \qquad V_{k+1}(x) = \inf_a (C + PV_k)(x,a), \quad x \in S.$$

Set

$$\lambda_k^- = \inf_x \{V_{k+1}(x) - V_k(x)\}, \qquad \lambda_k^+ = \sup_x \{V_{k+1}(x) - V_k(x)\}.$$

Then $\lambda_k^- \leq \lambda_k^+$. Often we will find $\lambda_k^+ - \lambda_k^- \to 0$ as $k \to \infty$.

Proposition 8.4

For all $k \geq 0$ and all controls u

$$\liminf_{n\to\infty} V_n^u(x)/n \geq \bar{\lambda}_k .$$

Moreover, if there exists $u : S \to A$ such that

$$V_{k+1}(x) = (c + PV_k)(x, u(x)), \quad x \in S,$$

then

$$\limsup_{n\to\infty} V_n^u(x)/n \leq \lambda_k^+ .$$

Proof Note that

$$\bar{\lambda}_k + V_k(x) \leq V_{k+1}(x) \leq (c + PV_k)(x, a), \quad x \in S, \ a \in A,$$
$$\lambda_k^+ + V_k(x) \geq V_{k+1}(x) = (c + PV_k)(x, u(x)), \quad x \in S$$

and apply Propositions 8.1 and 8.2. □