6 Dynamic optimization for non-negative costs

We show how to optimize a time-homogeneous stochastic controllable dynamical system with non-negative costs over an infinite time-horizon[19].

Let P be a time-homogeneous stochastic controllable dynamical system with state-space S and action-space A. Suppose given a *cost function*

$$c : S \times A \to \mathbb{R}^+.$$

Given a control u, define, as above, the *expected total cost function* V^u and the *infimal cost function* V by

$$V^u(x) = \mathbb{E}_x^u \sum_{n=0}^{\infty} c(X_n, U_n), \quad V(x) = \inf_u V^u(x).$$

Recall from Section 4 that $V_n^u(x) \uparrow V^u(x)$ as $n \to \infty$, where

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

Proposition 6.1. *Assume that A is finite. Then the infimal cost function is the minimal non-negative solution of the* dynamic 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 an optimal control, for every starting state x.

Proof. We know by Proposition 2.1 that V is a solution of the optimality equation. Suppose that F is another non-negative solution. We use the finiteness of A to find a map $\tilde{u} : S \to A$ such that

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

The argument leading to equation (2) is valid when $\beta = 1$, so we have

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

On letting $n \to \infty$, we obtain $F \geqslant V^{\tilde{u}} \geqslant V$. Finally, when $F = V$ we can take $\tilde{u} = u$ to see that $V \geqslant V^u$, and hence that u defines an optimal control. $\square$

The proposition allows us to see, in particular, that *value iteration* remains an effective way to approximate the infimal cost function in the current case. For let us set

$$V_n(x) = \inf_u V_n^u(x)$$

[19] This is also called negative programming – the problem can be recast in terms of non-positive rewards.

and note that $V_n(x) \uparrow V_\infty(x)$ as $n \to \infty$ for some function V_∞. Now $V_n^u \leqslant V^u$ for all n so, taking an infimum over controls we obtain $V_n \leqslant V$ and hence $V_\infty \leqslant V$. On the other hand we have the finite-horizon optimality equations

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

and we can pass to the limit as $n \to \infty$ to see that V_∞ satisfies the optimality equation. But V is the minimal non-negative solution of this equation, so $V_\infty \geqslant V$, so $V_\infty = V$.

A second iterative approach to optimality is the method of *policy improvement*. We know that, for any given map $u : S \to A$, we have

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

If V^u does not satisfy the optimality equation, then we can find a strictly better control by choosing $\tilde{u} : S \to A$ such that

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

with strict inequality at some state x_0. Then, obviously, $V^u \geqslant V_0^{\tilde{u}} = 0$. Suppose inductively that $V^u \geqslant V_n^{\tilde{u}}$. Then

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

so the induction proceeds and, letting $n \to \infty$, we obtain $V^u \geqslant V^{\tilde{u}}$, with strict inequality at x_0.