5 Dynamic optimization for discounted costs

We show how to optimize a time-homogeneous stochastic controllable dynamical system with bounded costs, discounted[18] at rate $\beta \in (0,1)$.

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

$$c : S \times A \to \mathbb{R},$$

and suppose that $|c(x,a)| \leqslant C$ for all x, a, for some constant $C < \infty$. Given a control u, define the *expected discounted cost function*

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

Define also the *infimal discounted cost function*

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

Our current set-up corresponds in the framework of Section 2, to the case of a time-dependent cost function $(n, x, a) \mapsto \beta^n c(x,a)$.

Define, for $n \geqslant 0$ and any control u,

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

Note that

$$|V_n^u(x) - V^u(x)| \leqslant C \sum_{k=n}^{\infty} \beta^k = \frac{C\beta^n}{1-\beta},$$

so, taking the infimum over u, we have

$$|V_n(x) - V(x)| \leqslant \frac{C\beta^n}{1-\beta} \to 0, \quad \text{as } n \to \infty.$$

[18]Such a discounting of future costs is normal in financial models, and reflects the fact that money can be invested to earn interest. There is a second way in which a discounted problem may arise. Consider the set-up of Section 4, modified by the introduction of a *killing time* T, with $\mathbb{P}(T \geqslant n+1) = \beta^n$ for all $n \geqslant 0$, independent of the controlled process $(X_n)_{n\geqslant 0}$. The idea is that, at each time step, independently, there is a probability β that some external event will terminate the process, and that no further rewards will be received. Then consider the *expected total reward function* for control u given by

$$V^u(x) = \mathbb{E}_x^u \sum_{n=0}^{T-1} r(X_n, U_n) = \mathbb{E}_x^u \sum_{n=0}^{\infty} r(X_n, U_n) 1_{\{T \geqslant n+1\}}.$$

Now

$$\mathbb{E}_x^u (r(X_n, U_n) 1_{\{T \geqslant n+1\}} | X_n, U_n) = \beta^n r(X_n, U_n),$$

so our problem reduces to the optimization of the *expected discounted reward function*

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

Taking advantage of time-homogeneity, the finite-horizon cost functions V_n may be determined iteratively for $n \geqslant 0$ by $V_0(x) = 0$ and the optimality equations

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

Hence, as in the case of non-negative rewards, we can compute V by *value iteration*.

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 an optimal control, for every starting state x.

Proof. We know that V satisfies the optimality equation by Proposition 2.1, and

$$|V(x)| \leqslant C \sum_{n=0}^{\infty} \beta^n = \frac{C}{1-\beta} < \infty,$$

so V is bounded. Let now F be any bounded solution of the optimality equation and let u be any control. Consider the process

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

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 $a \in A$,

$$\mathbb{E}_x^u(M_{n+1} - M_n | X_n = y, U_n = a) = \beta^n c(y, a) + \beta^{n+1} P F(y, a) - \beta^n F(y) \geqslant 0$$

and so

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

On letting $n \to \infty$, using the boundedness of F, we obtain $F \leqslant V^u$. Since u was arbitrary, this implies that $F \leqslant V$.

In the special case where we can find a stationary Markov control $u : S \to A$ such that

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

then, for all $y \in S$,

$$\mathbb{E}_x^u(M_{n+1} - M_n | X_n = y) = 0.$$

Hence

$$F(x) = \mathbb{E}_x^u(M_0) = \mathbb{E}_x^u(M_n) = V_n^u(x) + \beta^n \mathbb{E}_x^u F(X_n) \tag{2}$$

and so, letting $n \to \infty$, the final term vanishes and we find that $F = V^u$. In particular, in the case $F = V$, such a control u is optimal.

We do not know in general that there exists such a minimizing u but, given $\varepsilon > 0$, we can always choose $\tilde{u}$ such that

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

which we can write in the form

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

for a new cost function $\tilde{c} \geqslant c - \varepsilon$. The argument of the preceding paragraph, with $\tilde{c}$ in place of c and $\tilde{u}$ in place of u now shows that

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

Since $\varepsilon > 0$ was arbitrary, we conclude that $V = F$. $\square$