2 The dynamic optimality equation

We introduce a *cost function*

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

We assume throughout that one of the following three conditions holds: either c is non-negative, or c is non-positive, or there is a convergent series of constants $\sum_n c_n < \infty$ such that $|c(n, ., .)| \leqslant c_n$ for all n. In the second case, we shall usually express everything in terms of the *reward function* $r = -c$. Given a controllable dynamical system f and a control u, we define the *total cost function* $V^u : \mathbb{Z}^+ \times S \to \mathbb{R}$ by

$$V^u(k, x) = \sum_{n=k}^{\infty} c(n, x_n, u_n),$$

where $(x_n)_{n \geqslant k}$ is given by $x_k = x$ and $x_{n+1} = f(n, x_n, u_n)$ for all $n \geqslant k$. On the other hand, given a stochastic controllable dynamical system P with control u, we define the *expected total cost function* $V^u : \mathbb{Z}^+ \times S \to \mathbb{R}$ by

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

where $U_n = u_n(X_k, \ldots, X_n)$. In order to avoid the use of measure theory, we assume in the stochastic case, until Section 11, that the state-space S is countable. All the notions and results we present extend in a straightforward way to the case of a general measurable space $(S, \mathcal{S})$. Our assumptions on c are sufficient to ensure that the sums and expectations here are well-defined. The *infimal cost function* is defined by

$$V(k, x) = \inf_u V^u(k, x),$$

where the infimum is taken over all controls[9] [10]. A control u is *optimal* for (k, x) if $V^u(k, x) = V(k, x)$. The main problem considered in this course is the calculation of V and the identification of optimal controls u, when they exist, in this and some analogous

[9] Note that we have used a smaller class of controls in the deterministic case and a larger class in the stochastic case. Suppose we fix a starting time and state (k, x) and use a control u from the larger class in a deterministic controllable dynamical system, obtaining a controlled sequence $(x_n)_{n \geqslant k}$. Set $\tilde{u}_n = u_n(x_k, \ldots, x_n)$ for $n \geqslant k$, and define $\tilde{u}_n$ arbitrarily for $n \leqslant k - 1$. Then $\tilde{u}$ belongs to the smaller class of controls and has the same controlled sequence starting from (k, x). Hence there is no loss in restricting the infimum to the smaller class, but we should be mindful that the infimum can no longer be approached simultaneously for all (k, x) by a single sequence of controls. The larger class is necessary in the stochastic case to specify an appropriate dependence of the control on the controlled process.

[10] In some applications we shall have costs of the more general form $c(n, X_n, U_n, X_{n+1})$. However, we can always reduce to the case under discussion using the formula

$$\mathbb{E}^u_{(k,x)}(c(k, n_k, U_n, X_{n+1}) | X_k, \ldots, X_n) = \bar{c}(n, X_n, U_n),$$

where

$$\bar{c}(n, x, a) = \sum_{y \in S} P(n, x, a)_y c(n, x, a, y).$$

contexts. In most practical cases, a simple search over all controls is infeasible, because there are too many possibilities. Instead, the main approach is based on the following result.

Proposition 2.1. *The infimal cost function satisfies the* dynamic optimality equation[11]

$$\begin{aligned} V(k,x) &= \inf_a \{c(k,x,a) + V(k+1, f(k,x,a))\} \qquad (\textit{deterministic case}),\\ V(k,x) &= \inf_a (c+PV)(k,x,a) \qquad (\textit{stochastic case}), \end{aligned}$$

for all $k \geqslant 0$ and $x \in S$.

Proof. A simpler variant of the following argument may be used to prove the deterministic case. This is left as an exercise. Fix $k \in \mathbb{Z}^+$ and $x \in S$. Note that $V^u(k,x)$ depends on u only through $a = u_k(x)$ and through the control $\tilde{u}$, given for $n \geqslant k+1$ by $\tilde{u}_n(x_{k+1}, \dots, x_n) = u_n(x, x_{k+1}, \dots, x_n)$. By conditioning on X_{k+1}, we have

$$V^u(k,x) = c(k,x,a) + \sum_{y \in S} P(k,x,a)_y V^{\tilde{u}}(k+1,y).$$

Now, as we vary $\tilde{u}$ over all controls, we can approach the infimal value of $V^{\tilde{u}}(k+1,y)$ for all $y \in S$ simultaneously, so we obtain

$$\inf_{u_k(x)=a} V^u(k,x) = (c+PV)(k,x,a).$$

The optimality equation is now obtained on taking the infimum over $a \in A$. □

The idea of the proof is thus to condition on the first step and use the fact that the resulting constrained minimization is similar in form to the original. We used here the *Principle of Optimality.* In its most abstract form, this is just the fact that one can take an infimum over a set S given as a union $\cup_{a \in A} S_a$ by

$$\inf_{x \in S} f(x) = \inf_{a \in A} \inf_{x \in S_a} f(x).$$

Another, more concrete, instance is the fact any path of minimal length between two points must also minimize length between any two intermediate points on the path. Note that the proposition says nothing about uniqueness of solutions to the optimality equation. We shall look into this in a number of more specific contexts in the next few sections.

[11] Also called the *dynamic programming* or *Bellman* equation.