8 Dynamic optimization for long-run average costs

We show how to optimize the long-run average cost for a time-homogeneous stochastic controllable dynamical system with bounded instantaneous costs.

Let P be a time-homogeneous stochastic controllable dynamical system with state-space S and action-space A. Suppose given a bounded *cost function* $c : S \times A \to \mathbb{R}$. Define, as usual, for a control u,

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

where $U_k = u_k(X_0, \dots, X_k)$. A control u is *optimal*, starting from x, if the limit

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

exists and if, for all other controls $\tilde{u}$,

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

The limit λ is then 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*

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

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

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

Proof. Fix u and set

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

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 $a \in A$,

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

Hence

$$\theta(x) = \mathbb{E}_x^u(M_0) \leqslant \mathbb{E}_x^u(M_n) = \mathbb{E}_x^u(\theta(X_n)) - n\lambda + V_n^u(x)$$

and so

$$\frac{V_n^u(x)}{n} \geqslant \lambda + \frac{\theta(x)}{n} - \frac{\mathbb{E}_n^u(\theta(X_n))}{n}$$

and we conclude by letting $n \to \infty$. $\square$

By a similar argument, which is left as an exercise, one can also prove the following result.

Proposition 8.2. *Suppose there exists a constant λ and a bounded function θ on S, and a map $u : S \to A$, such that*

$$\lambda + \theta(x) \geqslant (c + P\theta)(x, u(x)), \quad x \in S.$$

Then, for all $x \in S$,

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

By combining the above two results, we see that, if λ and θ satisfy the *dynamic optimality equation*

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

and if the infimum is achieved at $u(x)$ for each $x \in S$, then λ is the minimal long-run average cost and u defines an optimal control, for all starting states x. Note that, since $P1 = 1$, we can add any constant to θ and still have a solution. So, we are free to impose the condition $\theta(x_0) = 0$ for any given $x_0 \in S$ when looking for solutions. The function θ can then be thought of as the (un-normalized) extra cost of starting at x rather than x_0.

Example (Consultant's job selection). Each day a consultant is either free or is occupied with some job, which may be of m different types $1, \dots, m$. Whenever he is free, he is given the opportunity to take on a job for the next day. A job of type x is offered with probability π_x and the types of jobs offered on different days are independent. On any day when he works 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?

We take as state-space the set $\{0, 1, \dots, m\}$, where 0 corresponds to the consultant being free and $1, \dots, m$ correspond to his working on a job of that type. The optimality equations for this problem are given by

$$\begin{aligned}
\lambda + \theta(0) =& \sum_{x=1}^{m} \pi_x \max\{\theta(0), \theta(x)\}, \\
\lambda + \theta(x) =& (1 - p_x)\theta(x) + p_x(R_x + \theta(0)), \quad x = 1, \dots, m.
\end{aligned}$$

Take $\theta(0) = 0$, then $\theta(x) = R_x - (\lambda/p_x)$ for $x = 1, \dots, m$, so the optimal λ must solve $\lambda = G(\lambda)$, where

$$G(\lambda) = \sum_{x=1}^{m} \pi_x \max\{0, R_x - (\lambda/p_x)\}.$$

Since G is non-increasing, there is a unique solution λ. The optimal control is then to accept jobs of type x if and only if $p_x R_x \geqslant \lambda$.

The optimality equation can be written down simply by reflecting on the details of the problem. A check on the validity of this process is provided by seeing how this particular problem can be expressed in terms of the general theory. For this, we take for state 0 the action-space $A_0 = \{(\varepsilon_1, \dots, \varepsilon_m) : \varepsilon_x \in \{0, 1\}\}$. Here the action $(\varepsilon_1, \dots, \varepsilon_m)$ signifies that we accept a job of type x if and only if $\varepsilon_x = 1$. There is no choice to be made in states $1, \dots, m$. We take, for $x = 1, \dots, m$,

$$P(0, \varepsilon)_x = \pi_x \varepsilon_x, \quad P(0, \varepsilon)_0 = \sum_{x=1}^{m} \pi_x (1 - \varepsilon_x), \quad P(x)_0 = p_x, \quad P(x)_x = 1 - p_x,$$

and

$$r(0,\varepsilon)=0,\quad r(x)=p_xR_x.$$

The reward function here gives the expected reward in state x, as in the discussion in footnote 10. We leave as an exercise to see that the general form of the optimality equations specializes to the particular equations claimed. The complicated form of action-space reflects the fact that, in this example, we in fact make our choice based on knowledge of the type of job offered, whereas, in the general theory, the action is chosen without such knowledge.

The following result provides a *value iteration* approach to long-run optimality. Recall that the finite-horizon optimality equations are $V_0(x)=0$ and, for $k\geqslant 0$,

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

Set

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

Proposition 8.3. *For all $k\geqslant 0$ and all controls u, we have*

$$\liminf_{n\to\infty}\frac{V_n^u(x)}{n}\geqslant\lambda_k^-.$$

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

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

Proof. Note that

$$\lambda_k^-+V_k(x)\leqslant V_{k+1}(x)\leqslant(c+PV_k)(x,a),\quad x\in S,\quad a\in A,$$

and

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

and apply the preceding two propositions with $\theta=V_k$. □