Discrete Dynamic Programming Notes

David K. Levine
March 11, 1998

We consider a decision problem taking place over time. In each time period, the single player can take an action by $\alpha \in A$, an action space. All information relevant to the future is incorporated in a state variable $y \in Y$, the state space. The dynamics of y are determined by a transition probability $\pi(y'|y,\alpha)$. We define the set of states reachable with some probability under some circumstances from a given state y as

$$S(y) \equiv \left\{ y' \middle| \exists \alpha \; \pi(y'|y,\alpha) > 0 \right\}.$$

We assume
(1) A is a compact subset of a finite dimensional space.
(2) Y is countable
(3) $\pi(y'|y,\alpha)$ is continuous in α
(4) $S(y)$ is finite

There are two main cases of interest

CASE 1: Y, A are finite

CASE 2: Y is a tree; only immediate successors have positive probability.

Preferences for this decision problem are given by a period utility $u(\alpha, y)$ and are additively separable over time (and states of nature) with discount factor $0 \le \delta < 1$. We assume

(5) u is bounded by $\bar{u}$ and continuous in α

Definitions

We denote finite histories by $h = (y_1, y_2, \ldots, y_t)$. For any given history, we may recover the length of the history $t(h) = t$, the final state in the history $y(h) = y_t$, the history through the previous period $h - 1 = (y_1, y_2, \ldots, y_{t-1})$, and the initial state $y_1(h)$. Histories are naturally ordered according to whether or not a history can logically follow from one another. We write $h' \ge h$. We say that a history is feasible if $y_\tau \in S(y_{\tau-1})$; the set of histories that is not feasible has probability zero. We denote by H the space of all feasible finite histories. Since we have assumed $S(y)$ finite, it follows that H is countable.

The object of choice is a strategy which is a map from histories to actions $\sigma : H \to A$. We denote by Σ the space of all strategies. A strategy is called *strong Markov* if $\sigma(h) = \sigma(h')$ if $y(h) = y(h')$; that is actions are determined entirely by the state. Any strong Markov strategy is equivalent to a map

$$\sigma : Y \to A.$$

Given a strategy we can define the probabilities of histories by

$$\pi(h | y_1, \sigma) \equiv$$
$$\begin{cases} \pi(y(h) | y(h-1), \sigma(h-1))\pi(h-1 | y_1, \sigma) & t(h) > 1 \\ 1 & t(h) = 1 \text{ and } y_1(h) = y_1 \\ 0 & t(h) = 1 \text{ and } y_1(h) \neq y_1 \end{cases}$$

We may also for any given initial state and strategy compute the expected average present value utility

$$V(y_1, \sigma) \equiv (1-\delta) \sum_{h \in H} \delta^{t(h)-1} u(\sigma(h), y(h)) \pi(h | y_1, \sigma).$$

The Dynamic Programming Problem

The problem which we call (*) is to maximize $V(y_1, \sigma)$ subject to $\sigma \in \Sigma$. A (not the) *value function* is any map $v : Y \to \Re$ bounded by $\bar{u}$.

Essential to the study of dynamic programming are two infinite dimensional objects: strategies and value functions. These naturally lie in two different spaces. Strategies naturally lie in $\Re^\infty$ the space of infinite sequences of numbers with the product topology. Value functions naturally lie in ℓ_∞ the space of bounded functions in the *sup* norm.

From the fact that the space of strategies is compact and utility continuous, it follows that

Lemma 1: a solution to (*) exists

This enables us to define *the* value function

$$v(y_1) \equiv \max_{\sigma \in \Sigma} V(y_1, \sigma)$$

The Bellman equation

We define a map $T:\ell_\infty \to \ell_\infty$ by $w' = T(w)$ if

$$w'(y_1) = \max_{\alpha\in A}(1-\delta)u(\alpha, y_1) + \delta \sum_{y'_1 \in S(y_1)} \pi(y'_1 | y_1, \alpha) w(y'_1).$$

We refer to the operator T as the Bellman operator.

Lemma 2: the value function is a fixed point of the Bellman equation $T(v) = v$

Lemma 3: the Bellman equation is a contraction mapping $\|T(w) - T(w')\| \le \delta \|w - w'\|$

Corollary: the Bellman equation has a unique solution

Conclusion 1: the unique solution to the Bellman equation is the value function

Lemma 4: there is a strong Markov optimum that may be found from the Bellman equation

proof: We define the strong Markov plan in the obvious way, show recursively that it yields a present value equal to the value function

$$\begin{aligned}
&v(y_1(h)) = \\
&(1-\delta)\textstyle\sum_{t(h)<T} \delta^{t(h)-1}\pi(h | y_1(h), \sigma) u(\sigma(h), y(h)) + \\
&(1-\delta)\textstyle\sum_{t(h)=T} \delta^{T}\pi(h | y_1(h), \sigma) v(h)
\end{aligned}$$