85-F-17

Path-Conserving and Path-Unimprovable Strategies in Dynamic Programming#*

by

Takao Kobayashi
The University of Tokyo

January 1986

PREPARED UNDER
THE PROJECT ON ECONOMIC THEORY
RESEARCH INSTITUTE FOR THE JAPANESE ECONOMY

ABSTRACT

The concepts of path-conserving and path-unimprovable strategies are proposed for a countable stage, countable state, dynamic programming problem where the objective is the maximization of the expectation of an arbitrary utility function defined on the sequence of states and actions. Conditions are provided for a path-conserving or a path-unimprovable strategy to be optimal. Relationship to the findings of Strauch, Blackwell and Kreps are also considered.

1. Introduction.

Given a strategy[1/] of a dynamic programming problem one can ask if it satisfies the following two similar, but different questions at each decision point. One is: Provided that he must use the optimal strategy from tomorrow and thereafter, can he achieve the overall optimal payoff if he follows the given strategy today? The other is: Provided that he must use the given strategy from tomorrow and thereafter, is it best for him to follow this strategy today? If the answer to the first question is yes, the strategy is said to be conserving. If the answer to the second question is yes, the strategy is said to be unimprovable.

Clearly an optimal strategy is both conserving and unimprovable. The interesting question is the converse.

It is known that if the problem is of a finite time-horizon any conserving or any unimprovable strategy is optimal.[2/] However, these converse statements do not always hold for infinite-horizon problems.

The following counterexample will help to understand the problem. Consider a Markov decision problem with two states a and b. State b is "absorbing"; that is, if the system once reaches state b it stays there forever. Two actions m ("move") and s ("stay") are available when the system is in state a. If action m is taken the system moves to state b; if action s is taken the system stays in state a until next stage. Reward of $1 accrues from the former action and no reward accrues from the latter action. The time horizon is infinite, and the decision maker is interested in maximizing the no-discounted sum of rewards. Figure 1 below illustrates the structure of the problem.[3/]

Fig. 1

Clearly the optimal strategy is to choose m in state a. If the initial state is a, it yields total reward of \$1. If the initial state is b, total reward is \$0 regardless of the strategy.

Let π_s be a strategy that always chooses s in state a. This strategy is conserving, because if the state is a today, by using π_s today and then using the optimal strategy from tomorrow he expects the maximum total reward of \$1. However, if he uses strategy π_s throughout the time horizon his total reward is \$0. So π_s is not an optimal strategy. Thus we have a case in which a conserving strategy is not optimal.

A standard procedure of constructing an optimal strategy is to find at each date and in each state an action which maximizes the total expected payoff given that one will follow the optimal path from tomorrow. Bellman's principle of optimality underlies such a procedure. However, the above example shows that it does not necessarily generate an optimal strategy.

Strauch [1966] proved that any conserving strategy is optimal and hence the standard procedure is legitimate in negative dynamic programming problems. But the above example is a case of positive dynamic programming and is not covered by his theorem.

Next, assume that action m incurs a cost of \$1 instead of reward. This time the decision maker is interested in minimizing the no-discounted sum of costs. In this case the optimal strategy is to choose s in state a, and the minimum total cost is \$0 regardless of the initial state.

Let π_m be a strategy that chooses m in state a. This strategy is unimprovable, because if the state is a today and if he must use π_m from tomorrow, he has no incentive to deviate from π_m today. In other words, it does not save any overall cost to choose s in state a if he must use π_m from the next stage. However, if he uses strategy π_m throughout the time horizon his total cost is $1. So $\pi_{\underline{m}}$, which is unimprovable, is not an optimal strategy.

Blackwell [1967] proved that any unimprovable strategy is optimal in positive dynamic programming. But since the current problem is a case of negative dynamic programming, it is not covered by his theorem. Later, Strauch's and Blackwell's theorems were unified by Kreps [1977] with the notion of upper- and lower-convergent dynamic programming.[4/]

In this paper we propose a somewhat different set of definitions for conserving and unimprovable strategies. It is sometimes too much to require these properties on a strategy at every decision point. Associated with any strategy is a set of paths which are realizable. And one may only be interested to see if the properties are fulfilled along these paths. We define that a strategy is path-conserving if it is conserving at all decision points that can be reached with positive probabilities. Similarly, we define that a strategy is path-unimprovable if it is unimprovable at all decision points that can be reached with positive probabilities.

This paper establishes the following. For a path-conserving strategy to be optimal one only needs Kreps' upper convergence of the utility function. Thus any path-conserving strategy is optimal in negative dynamic programming, or more in general, in upper-convergent dynamic programming. On the other hand, a path-unimprovable strategy may not be optimal in positive, or lower-convergent dynamic programming. It may not be optimal even in finite-horizon problems. We will show that the additional conditions required are convexity and differentiability.

Section 2 gives the general formulation of the problem. Section 3 defines the notions of a path-conserving strategy and a path-unimprovable strategy. We also provide a simple finite-horizon problem in which a path-unimprovable strategy is not optimal. Section 4 gives some preliminary propositions which hold to the extent that the utility function is lower- or upper-convergent. Section 5 provides the main proposition. Section 6 summarizes the results. It also includes a brief description of an application of our theorem to a problem in intertemporal capital markets, where Strauch-Blackwell-Kreps' version of conserving and unimprovable strategies fails to apply.

2. Formulation of the Problem.

To secure a maximum generality of the problem we use Kreps' formulation of a dynamic programming problem with a minor modification. It consists of[5/]:

(a) countable <u>state spaces</u> X_t with generic x_t, for $t = 1, 2, \ldots$;

(b) a countable <u>initial history space</u> H_0 with generic h_0;

(c) <u>action spaces</u> $A_t(h_t)$ with generic a_t, for $h_t \in H_t$ and $t = 0, 1, 2, \ldots$, in which

(d) H_t, for $t = 1, 2, \ldots$, are <u>partial history spaces</u> with generic h_t, defined iteratively by $h_t := (h_{t-1}, a_{t-1}, x_t)$, or $h_t = (h_0, a_0, x_1, a_1, x_2, \ldots, a_{t-1}, x_t)$;

(e) <u>transition probabilities</u> $P_t(x_{t+1}|a_t, h_t)$ for $a_t \in A_t(h_t)$, $h_t \in H_t$, $x_{t+1} \in X_{t+1}$, $t = 0, 1, 2, \ldots$, such that $P_t(x_{t+1}|a_t, h_t) \geq 0$ and $\Sigma_{x_{t+1}} P_t(x_{t+1}|a_t, h_t) = 1$; and

(f) an extended real valued <u>utility function</u> U defined on the space of <u>complete histories</u> H with generic $h = (h_0, a_0, x_1, a_1, x_2, \ldots)$.

A *strategy* is a complete description of choices of actions at all partial histories. Thus the *strategy space* is defined by $\Pi := X_{t=0}^{\infty} X_{h_t} A_t(h_t)$, with generic π. We write $\pi(h_t)$ as the projection of π onto $A_t(h_t)$. For each π and h_t, *conditional* probabilities $P^{\pi}(\cdot|h_t)$ and *conditional expectations* $E^{\pi}[\cdot|h_t]$ *using* π *given* h_t are constructed on H from the transition probabilities in the usual fashion. *Unconditional probabilities* $P^{\pi}(\cdot)$ *using* π are constructed in the same spirit.

For each t and h_t, the *expected utility using* π *given* h_t is defined by[6/]

$$v_t(\pi, h_t) := E^{\pi}[U(h)|h_t],$$

and the *optimal expected utility given* h_t is defined by

$$f_t(h_t) := \sup_{\pi \in \Pi} v_t(\pi, h_t).$$

Note that $v_t(\pi,h_t)$ is the *entire* expected utility. It is not the expected *additional* utility accruing after time t, which is more conventional in the dynamic programming literature but only makes sense when U is time-additive.

3. Path-Optimal, Path-Conserving, and Path-Unimprovable Strategies.

Given a strategy π let $\mathcal{R}(\pi)$ denote the set of all π-*reacheable* partial histories, which is defined by

$$\mathcal{R}(\pi) := \bigcup_{t=0}^{\infty} \{h_t | P^{\pi}(h_t) > 0\}.$$

We say that strategy π is *optimal at* h_t if

$$v_t(\pi, h_t) = f_t(h_t). \qquad (3.1)$$

Eq.(3.1) implies that using π at time t and thereafter yields the optimal expected utility given h_t. We also say that strategy π is *path-optimal* if it is optimal at every $h_t \in \mathcal{R}(\pi)$.

For π to be an optimal strategy for the entire problem with initial history h_0, it must be that $v_0(\pi, h_0) = f_0(h_0)$. It is easy to show that this requires that π be optimal at every π-reacheable partial history, i.e., π be path-optimal.[7/] To show this equivalence (and for later purposes) the following lemma is useful.

LEMMA 1. For all t and t', with $t' > t$, and all $\pi \in \Pi$, $h_t \in H_t$,

$$v_t(\pi, h_t) = E^{\pi}[v_{t'}(\pi, h_{t'}) | h_t].$$

PROOF. From the definition of $v_t(\pi, h_t)$ and the law of conditional expectation it follows that

$$\begin{aligned} v_t(\pi, h_t) &= E^{\pi}[U(h) | h_t] \\ &= E^{\pi}[E^{\pi}[U(h) | h_{t'}] | h_t] \\ &= E^{\pi}[v(\pi, h_{t'}) | h_t], \end{aligned}$$

since $h_{t'}$ "contains" h_t. ■

PROPOSITION 1. For any $h_0 \in H_0$, a strategy π satisfies

$$v_0(\pi, h_0) = f_0(h_0)$$

if and only if π is path-optimal.

PROOF. The "if" part directly follows from the definition of path-optimality. To show the "only if" part, assume the contrary. Then there exist t and $\hat{h}_t \in \mathcal{R}(\pi)$ for which

$$v_t(\pi, \hat{h}_t) < f_t(\hat{h}_t) = \sup_{\pi'} v_t(\pi', \hat{h}_t).$$

So we can find $\hat{\pi} \in \Pi$ such that

$$v_t(\pi, \hat{h}_t) < v_t(\hat{\pi}, \hat{h}_t).$$

Let π' be a strategy which uses π until time t, and follows $\hat{\pi}$ afterwards if $\hat{h}_t$ is reached at time t, and otherwise follows π. Using lemma 1 we get

$$v_0(\pi', h_0) = E^{\pi'}[v_t(\pi', h_t)|h_0]$$
$$= E^{\pi}[v_t(\pi', h_t)|h_0].$$

Noting that $v_t(\pi', h_t) = v_t(\hat{\pi}, h_t) > v_t(\pi, h_t)$ if $h_t = \hat{h}_t$ and $v_t(\pi', h_t) = v_t(\pi, h_t)$ otherwise and that $P^{\pi}(\hat{h}_t) > 0$, we obtain

$$v_0(\pi', h_0) > E^{\pi}[v_t(\pi, h_t)|h_0]$$
$$= v_0(\pi, h_0).$$

This contradicts the assumption that $v_0(\pi, h_0) = f_0(h_0)$. ■

Now we give definitions of a path-conserving and a path-unimprovable strategy.

DEFINITION 1. For any t and $h_t \in H_t$, a strategy π is said to be <u>conserving at</u> h_t if it satisfies

$$f_t(h_t) = E^{\pi}[f_{t+1}(h_{t+1})|h_t]. \tag{3.2}$$

A strategy π is said to be <u>path-conserving</u> if it is conserving at every $h_t \in \mathcal{R}(\pi)$.

The right-hand-side of (3.2) is the conditional expectation of the optimal expected utility $f_{t+1}(h_{t+1})$ given that the partial history until time t is h_t and that an action is taken at t according to strategy π. This equals the expected utility given h_t of using π at t and following the optimal path afterwards. A strategy is conserving at h_t if this expected utility equals the optimal expected utility given h_t.

DEFINITION 2. For any t and $h_t \in H_t$, a strategy π is said to be <u>unimprovable at</u> h_t if it satisfies

$$v_t(\pi, h_t) = \sup_{\pi'} E^{\pi'}[v_{t+1}(\pi, h_{t+1}) | h_t]. \qquad (3.3)$$

A strategy π is said to be path-unimprovable if it is unimprovable at every $h_t \in \mathcal{R}(\pi)$.

The right-hand-side of (3.3) is the supremum of the expected utilities given that the partial history until time t is h_t and that strategy π is used at and after time t+1, where supremum is taken with respect to the action at t. Thus a strategy is unimprovable at h_t if it cannot be improved upon by deviating from the strategy only at time t.

It is straightforward to establish that any path-optimal strategy is both path-conserving and path-unimprovable.

PROPOSITION 2. If a strategy π is path-optimal, then (a)it is path-conserving, and (b)it is path-unimprovable.

PROOF. (a) If π is path-optimal, for any t and $h_t \in \mathcal{R}(\pi)$ we have

$$f_t(h_t) = v_t(\pi, h_t).$$

From this and lemma 1 we obtain

$$f_t(h_t) = E^{\pi}[v_{t+1}(\pi, h_{t+1}) | h_t],$$
$$= E^{\pi}[f_{t+1}(h_{t+1}) | h_t],$$

where the second equality follows from the condition that $v_{t+1}(\pi, h_{t+1}) = f_{t+1}(h_{t+1})$ for all $h_{t+1} \in \mathcal{R}(\pi)$.

(b) Suppose that π is not path-unimprovable. Then there exist t, $\hat{h}_t \in \mathcal{R}(\pi)$, and $\hat{\pi}$ such that

$$v_t(\pi, \hat{h}_t) < E^{\hat{\pi}}[v_{t+1}(\pi, h_{t+1}) | \hat{h}_t].$$

The right-hand-side is the expected utility given $\hat{h}_t$ of a strategy π' which chooses an action at time t according to $\hat{\pi}$ if the partial history is $\hat{h}_t$ and uses strategy π otherwise. Therefore, we obtain $v_t(\pi, \hat{h}_t) < v_t(\pi', \hat{h}_t)$. This contradicts the assumption that π is path-optimal. ■

The purpose of the rest of this paper is to investigate the conditions under which the converse statements are true. As shown in the following example a path-unimprovable strategy may not be path-optimal even in finite-horizon problems.

EXAMPLE. Fig. 2 below describes the problem in the form of a

Fig. 2

decision tree. One starts at the initial "node" h_0, where he chooses between two actions —— to take the upper branch or to take the lower branch (thus, the transition probabilities are degenerate). At each node, h_1' or h_1'', he again chooses between two branches. Thus there are four alternative paths. The numbers in the right-end are the utilities associated with these paths. Let π denote the strategy indicated by the arrows. Obviously this strategy is not path-optimal, since $f_0(h_0) = 5$. However, it is unimprovable at h_1'. More importantly, it is unimprovable at h_0, since if he must take the lower branch at h_1'' his best choice at h_0 is to go to h_1'. Therefore the strategy π is path-unimprovable.

4. Sequentially Path-Conserving and Sequentially Path-Unimprovable Strategies.

We now introduce another set of notions, sequentially path-conserving strategies and sequentially path-unimprovable strategies. These notions are stronger than their counterparts of the previous section.

DEFINITION 3. For any t and $h_t \in H_t$, a strategy π is said to be <u>sequentially conserving at</u> h_t if it satisfies

$$f_t(h_t) = E^{\pi}[f_{t+k}(h_{t+k})|h_t] \tag{4.1}$$

for any positive integer k. A strategy π is said to be <u>sequentially path-conserving</u> if it is sequentially conserving at every $h_t \in \mathcal{R}(\pi)$.

A path-conserving strategy satisfies (4.1) for $k = 1$. Thus the notion of sequentially path-conserving strategies extends the idea of (one-step) path-conserving strategies to an arbitrary number of steps. Similar construction is made for the notion of path-unimprovable strategies.

DEFINITION 4. For any t and $h_t \in H_t$, a strategy π is said to be <u>sequentially unimprovable at</u> h_t if it satisfies

$$v_t(\pi, h_t) = \sup_{\pi'} E^{\pi'}[v_{t+k}(\pi, h_{t+k})|h_t] \tag{4.2}$$

for any positive integer k. A strategy π is said to be <u>sequentially path-unimprovable</u> if it is sequentially unimprovable at every $h_t \in \mathcal{R}(\pi)$.

Let us investigate the relationship between path-optimal strategies and sequentially path-conserving (sequentially path-unimprovable) strategies. First, it is straightforward to see that any path-optimal strategy is both sequentially path-conserving and sequentially path-unimprovable.[8/] Further, in finite-horizon problems any sequentially path-conserving (sequentially path-unimprovable) strategy is path-optimal, which one can easily verify by taking k larger than the time-horizon. In the following we show that Strauch-Kreps' condition (Blackwell-Kreps' condition) is exactly what is needed for a sequentially path-conserving (sequentially path-unimprovable) strategy to be path-optimal.

UPPER- AND LOWER CONVERGENT UTILITY FUNCTIONS.

For T = 0, 1, 2, ... define $\bar{U}^T$ and $\underline{U}^T$ on H_T by

$$\bar{U}^T(h_T) := \sup\{U(h) | h \in H,\ h_T(h) = h_T\},$$

and

$$\underline{U}_T(h_T) := \inf\{U(h) | h \in H,\ h_T(h) = h_T\},$$

where $h_T(h)$ denotes the projection of h from H to H_T. Given a partial history up to time T, h_T, $\bar{U}^T(h_T)$ measures the overall utility with the most "optimistic" estimate of the subsequent path post time T. $\underline{U}^T(h_T)$ corresponds to the most "pessimistic" estimate. Obviously, for all $h \in H$,

$$\bar{U}^0(h) \geq \bar{U}^1(h) \geq \cdot \cdot \cdot \geq U(h), \tag{4.3}$$

and

$$\underline{U}^0(h) \leq \underline{U}^1(h) \leq \cdot \cdot \cdot \leq U(h), \tag{4.4}$$

where $\bar{U}^T(h)$ and $\underline{U}^T(h)$ stand for $\bar{U}^T(h_T(h))$ and $\underline{U}^T(h_T(h))$, respectively.

DEFINITION 5. A utility function U is <u>upper convergent</u> if $\bar{U}^0(h) < +\infty$ and $\lim_{T\to\infty}\bar{U}^T(h) = U(h)$ for all $h \in H$.

DEFINITION 6. A utility function U is <u>lower convergent</u> if $\underline{U}^0(h) > -\infty$ and $\lim_{T\to\infty}\underline{U}^T(h) = U(h)$ for all $h \in H$.

DEFINITION 7. A utility function U is <u>convergent</u> if it is both upper- and lower convergent.

The utility function of a finite-horizon problem is trivially convergent. Also, if the utility function is given by $U(h) = \Sigma_{t=0}^{\infty} r_t(x_t, a_t)$ with $r_t(x_t, a_t) \leq 0$ (the classic case of <u>negative dynamic programming</u>), it is upper convergent. If U(h) is of the same form with $r_t(x_t, a_t) \geq 0$ (the classic case of <u>positive dynamic programming</u>) it is lower convergent.[9/]

Given a utility function $\bar{U}^T$ defined on H_T one can construct a T-horizon dynamic programming subproblem; namely, the problem of selecting actions contingent on histories at times t = 0, 1, .., T-1, so as to maximize the expectation of $\bar{U}^T$. Any strategy $\pi \in \Pi$ for the overall problem can be restricted to $\times_{t=0}^{T-1} (\times_{h_t} A_t(h_t))$ to serve as a strategy for the T-horizon subproblem. Let $\bar{v}_t^T(\pi, h_t)$ denote the expected utility using π given h_t of such a subproblem, and let $\bar{f}_t^T(h_t)$ denote the optimal expected utility given h_t.

A similar construction is possible for the utility function $\underline{U}^T$. Define the expected utility $\underline{v}_t^T(\pi, h_t)$ and the optimal expected utility $\underline{f}_t^T(h_t)$, correspondingly.

Kreps [1977] proved the following.

(a) If the utility function U is upper convergent, then

$$\lim_{T\to\infty}\bar{v}_t^T(\pi, h_t) = v_t(\pi, h_t) \text{ and } \lim_{T\to\infty}\bar{f}_t^T(h_t) = f_t(h_t)$$

for all π, t and h_t;

(b) If the utility function U is lower convergent, then

$$\lim_{T\to\infty} \underline{v}_t^T(\pi, h_t) = v_t(\pi, h_t) \text{ and } \lim_{T\to\infty} \underline{f}_t^T(h_t) = f_t(h_t)$$

for all π, t and h_t.

These results endorse the computational procedure known as value iteration. We utilize these limit theorems to obtain the following two propositions.

PROPOSITION 3. If the utility function is upper convergent, any sequentially path-conserving strategy is path-optimal.

PROOF. For any $\pi \in \Pi$, $t \geq 0$, $T > t$ and $h_t \in H_t$, we have

$$\bar{v}_t^T(\pi, h_t) = E^{\pi}[\bar{U}^T(h_T) | h_t]$$
$$\geq E^{\pi}[f_T(h_T) | h_t],$$

since $\bar{U}^T(h_T) \geq f_T(h_T)$ by (4.3). If π is sequentially path-conserving, for all $h_t \in \mathcal{R}(\pi)$

$$f_t(h_t) = E^{\pi}[f_T(h_T) | h_t].$$

Thus we obtain

$$\bar{v}_t^T(\pi, h_t) \geq f_t(h_t).$$

Letting $T \to \infty$ and using (a) above yields

$$v_t(\pi, h_t) \geq f_t(h_t)$$

for all $h_t \in \mathcal{R}(\pi)$. This implies that π is path-optimal. ■

PROPOSITION 4. If the utility function U is lower convergent, any sequentially path-unimprovable strategy is path-optimal.

PROOF. If π is sequentially path-unimprovable, for any $t \geq 0$, $k \geq 1$ and $h_t \in \mathcal{R}(\pi)$ we have

$$v_t(\pi, h_t) = \sup_{\pi'} E^{\pi'}[v_{t+k}(\pi, h_{t+k}) | h_t]$$
$$\geq \sup_{\pi'} E^{\pi'}[\underline{U}^{t+k}(h_{t+k}) | h_t],$$

where the inequality follows from (4.4). For the (t+k)-horizon subproblem with the utility function $\underline{U}^{t+k}$, the k-step optimality equation

$$\underline{f}_t^{t+k}(h_t) = \sup_{\pi'} E^{\pi'}[\underline{f}_{t+k}^{t+k}(h_{t+k})|h_t]$$

holds.[10/] Since $\underline{f}_{t+k}^{t+k}(h_{t+k}) = \underline{U}^{t+k}(h_{t+k})$ by definition, combining these two relations yields

$$v_t(\pi, h_t) \geq \underline{f}_t^{t+k}(h_t).$$

If we let $k \to \infty$ and use (b) above, we obtain

$$v_t(\pi, h_t) \geq f_t(h_t)$$

for all $h_t \in \mathcal{R}(\pi)$. Hence π is path-optimal. ■

5. Optimality of Path-Conserving and Path-Unimprovable Strategies.

Thus far we have shown that upper (lower) convergence of the utility function is sufficient to guarantee that any sequentially path-conserving (sequentially path-unimprovable) strategy is path-optimal. The work left is to investigate an appropriate condition under which any path-conserving (path-unimprovable) strategy is sequentially path-conserving (sequentially path-unimprovable).

We first show that a path-conserving strategy is always sequentially path-conserving.

PROPOSITION 5. Any path-conserving strategy is sequentially path-conserving.

PROOF. If π is path-conserving, for any $t \geq 0$ and $h_t \in \mathcal{R}(\pi)$ we have

$$f_t(h_t) = E^{\pi}[f_{t+1}(h_{t+1})|h_t].$$

Similarly, for any $h_{t+1} \in \mathcal{R}(\pi)$ we have

$$f_{t+1}(h_{t+1}) = E^{\pi}[f_{t+2}(h_{t+2})|h_{t+1}].$$

Since the expected value in the right-hand-side of the first equation does not depend on $f_{t+1}(h'_{t+1})$ for $h'_{t+1} \notin \mathcal{R}(\pi)$, we can substitute the second equation into the right-hand-side of the first equation. Using lemma 1 we obtain

$$f_t(h_t) = E^{\pi}[E^{\pi}[f_{t+2}(h_{t+2})|h_{t+1}]|h_t]$$
$$= E^{\pi}[f_{t+2}(h_{t+2})|h_t].$$

By repeating the above procedure (or, by induction), we get

$$f_t(h_t) = E^{\pi}[f_{t+k}(h_{t+k})|h_t]$$

for $t \geqq 0$, $h_t \in \mathcal{R}(\pi)$, and any positive integer k. Hence, π is sequentially path-conserving. ■

COROLLARY. <u>If the utility function is upper convergent, any path-conserving strategy is path-optimal.</u>

We now ask whether any path-unimprovable strategy is sequentially path-unimprovable. It turns out that convexity and differentiability of the problem is essential for this statement to be true. The following lemma plays a key role to establish our result.

LEMMA 2. <u>Let Y be a locally convex, linear topological space and A a convex set in Y. Let $F: A \to R$ and $G: A \to R$ be concave functions such that $G(a) \geqq F(a)$ for all $a \in A$ with equality holding at $a = a^{\circ} \in icr(A)$.</u>[11/] <u>Further we assume that the gradient</u>[12/] <u>$\nabla F(a^{\circ})$ exists. Then, if F attains its maximum at a°, G also attains its maximum at a°.</u>[13/]

PROOF. Since F is concave on A and continuous at a°, a theorem due to Pshenichnii[14/] entails that there exists a <u>supergradient</u>[15/] of F at a°, denoted by Φ, such that $\Phi(a^{\circ}) = \max\{\Phi(a)|a \in A\}$. If F is dominated by G with $F(a^{\circ}) = G(a^{\circ})$, the set of supergradients of F at a°, which is usually called the superdifferential of F at a° and denoted by $\partial F(a^\circ)$, contains the superdifferential $\partial G(a^\circ)$ of G at a°, since $\Psi(a - a^\circ) \geqq G(a) - G(a^\circ)$ implies $\Psi(a - a^\circ) \geqq F(a) - F(a^\circ)$ for an arbitrary Ψ. On the other hand, if the gradient $\nabla F(a^\circ)$ exists in the dual space of Y the superdifferential $\partial F(a^\circ)$ must consist of a single element, namely $\nabla F(a^\circ)$, and if $a^\circ \in icr(A)$ the superdifferential $\partial G(a^\circ)$ must be nonempty.[16/] Therefore Φ must also be a supergradient of G at a°, namely, $\Phi \in \partial G(a^\circ)$. Applying Pshenichnii's theorem again it follows that $G(a^\circ) = \max \{G(a) | a \in A\}$. ■

We impose the following assumption to apply lemma 2 to our problem.

ASSUMPTION X. For each $t \geqq 0$ and $h_t \in H_t$ let the action space $A_t(h_t)$ be a convex set in a locally convex, linear topological space. Further, assume that for any $\pi \in \Pi$ the function which maps each $a_t \in A_t(h_t)$ to $E^{a_t}[v_{t+1}(\pi, h_{t+1}) | h_t]$ is a concave function on $A_t(h_t)$ and its gradient exists at any point in $icr(A_t(h_t))$.

PROPOSITION 6. Under assumption X, any path-unimprovable strategy $\pi \in \Pi$ is sequentially path-unimprovable if $\pi(h_t) \in icr(A_t(h_t))$ for each $t \geqq 0$ and $h_t \in \mathcal{R}(\pi)$.

PROOF. For each $t \geqq 0$ and $h_t \in \mathcal{R}(\pi)$, define functions F_t and G_t on $A_t(h_t)$ by

$$F_t(a_t) = E^{a_t}[v_{t+1}(\pi, h_{t+1}) | h_t]$$

and

$$G_t(a_t) = E^{a_t}[\sup_{\pi'} E^{\pi'}[v_{t+2}(\pi, h_{t+2}) | h_{t+1}] | h_t].$$

Since π is path-unimprovable we have

$$F_t(\pi(h_t)) = \max\ \{F_t(a_t) | a_t \in A_t(h_t)\},$$

and also

$$G_t(\pi(h_t)) = E^{\pi(h_t)}[\sup_{\pi'} E^{\pi'}[v_{t+2}(\pi,\ h_{t+2})|h_{t+1}]|h_t]$$
$$= E^{\pi(h_t)}[E^{\pi(h_{t+1})}[v_{t+2}(\pi,\ h_{t+2})|h_{t+1}]|h_t]$$
$$= E^{\pi}[v_{t+1}(\pi,\ h_{t+1})|h_t]$$
$$= F_t(\pi(h_t)).$$

Further, we clearly have

$$G_t(a_t) \geqq F_t(a_t) \text{ for all } a_t \in A_t(h_t).$$

Hence we can apply lemma 2 to obtain

$$G_t(\pi(h_t)) = \max\{G_t(a_t)|a_t \in A_t\}.$$

Since $G_t(\pi(h_t)) = F_t(\pi(h_t)) = v_t(\pi,\ h_t)$, this implies

$$v_t(\pi,\ h_t) \geqq G_t(a_t) \text{ for all } a_t \in A_t(h_t),$$

or equivalently,

$$v_t(\pi,\ h_t) = \sup_{\pi'} E^{\pi'}[v_{t+2}(\pi,\ h_{t+2})|h_t].$$

By repeating the above procedure (or, by induction), we obtain for all $t \geqq 0$, $h_t \in \mathcal{R}(\pi)$ and $k \geqq 1$,

$$v_t(\pi,\ h_t) = \sup_{\pi'} E^{\pi'}[v_{t+k}(\pi,\ h_{t+k})|h_t].$$

Hence π is sequentially path-unimprovable. ■

COROLLARY. <u>If the utility function U is lower convergent and assumption X is satisfied, any path-unimprovable strategy π with $\pi(h_t) \in \text{icr}(A_t(h_t))$ for any $t \geqq 0$ and $h_t \in \mathcal{R}(\pi)$ is path-optimal.</u>

It is important to note that this proposition requires that the strategy π must at each moment choose an action which is in the "interior" of the action space. If we go back to the example of section 3, any "pure" strategy, including the one we focused on, chose actions on the boundary of the action space, because to "convexify" the problem the action space at each node must be defined as the one-dimensional simplex $\{(\alpha_1, \alpha_2) | \alpha_1 + \alpha_2 = 1, \alpha_1 \geq 0, \alpha_2 \geq 0\}$, where α_1 and α_2 are the probabilities of a mixed strategy.

6. Summary.

What we obtained in this paper are summarized in Fig. 3 below. Arrows with solid lines indicate that no assumption is needed to move in the

Fig. 3

pointed direction. Directions which necessitates some assumptions are indicated by arrows with dotted lines. The symbol "u.c." ("l.c.") stands for upper (lower) convergence of the utility function.

If we omit the term "path-" and add "at every partial history" in each box, we revert to the problem dealt by Kreps. In that case the broken arrow with "Assumption X & interior strategy" becomes a solid arrow, and no other part of the diagram changes. Thus the last point is really the difference between our construction and Strauch-Blackwell-Kreps'. Our notions of sequentially conserving and sequentially unimprovable strategies helped to clarify this difference.

The results of this paper have an interesting economic application. Here we briefly outline the problem. Consider two separate exchange economies, economy 1 and economy 2, each consisting of a single "representative" consumer/investor. Each economy has markets which provide opportunities to reallocate the individual's commodity endowments. Financial assets are also available for reallocating "purchasing power" over time. These two economies are identical in the individual's endowments, preferences, the economies' productions and the available markets. They are different only on one aspect. Namely, the individual of economy 1 is very naive; when he constructs a portfolio of the financial assets he never considers possibilities of reselling in the future. In other words, he is a pure "investor", who purchases financial assets merely to "eat up" dividends and interests.[17/] In contrast, the individual of economy 2 is more sophisticated and may act as a "speculator" who buys and sells financial assets to realize capital gains or to avoid capital losses. What we are concerned with is the equilibrium prices of the financial assets in these two economies and their mutual relationship.

The equilibrium price systems of these economies can be analyzed using the notions of a path-conserving and a path-unimprovable strategy. Let π^h denote the "buy-and-hold" strategy, which at each time and on every contingency instructs the individual not to retrade any financial asset. It is not hard to understand the statement: (a) A price system is an equilibrium of economy 1 if and only if π^h is a path-unimprovable strategy provided that this price system prevails over time; (b) A price system is an (rational-expectations) equilibrium of economy 2 if and only if π^h is a path-conserving strategy provided that this price system prevails over time. One can use the propositions of this paper to relate these two equilibrium price systems. For a formal treatment of the subject the readers are led to a companion paper, Kobayashi [1986]. It also explains why the notions of Strauch-Blackwell-Kreps are inappropriate to analyze this problem.

* This paper originates from a theorem proved in Kobayashi [1983]. In a private conversation David Kreps taught me a possible link between the theorem and his theorem on conserving and unimprovable strategies in dynamic programming. The author wishes to gratefully acknowledge his suggestion, without which he would never have thought of the present, extensively generalized version of the theorem. The preparation of this paper was partially supported by a research grant from the Kikawada Foundation.

1/ A strategy specifies an action at every decision point.

2/ This result will become evident in the sequel.

3/ This example is due to Ross [1974].

4/ The notions of negative, positive, upper-convergent, and lower-convergent dynamic programming will be defined in section 4.

5/ The nonessential difference between this and Kreps' formulation is that the latter defines partial histories as $h_t := (h_0, x_1, \ldots, x_t)$, allowing the state variable x_t to include a_{t-1} when history records previous actions. We choose to record past actions explicitly in h_t, because we want countable state spaces and convex (uncountable) action spaces to coexist.

6/ A suitable assumption must be made to ensure the existence of the integrals. Kreps assumes that for each $h_0 \in H_0$, U(h) is bounded either above or below uniformly in h for which the initial history is h_0.

7/ Strauch-Blackwell-Kreps' definition of an optimal strategy requires that the strategy be optimal at every $h_t \in H_t$. But, whether a strategy π is optimal at π-unreacheable partial histories h_t (i.e., $P^{\pi}(h_t) = 0$) is irrelevant to the overall optimality of π given an initial history.

8/ The proof is essentially the same as that of proposition 2.

9/ Kreps lists additional examples, e.g., $U(h) := \inf\{r_1(x_1), r_2(x_2), \ldots\}$ is upper convergent, and $U(h) := \sup\{r_1(x_1), r_2(x_2), \ldots\}$ is lower convergent.

10/ For a proof of the one-step version, see Kreps. The k-step version is a trivial extension of the one-step version.

11/ The symbol "icr(A)" is the <u>intrinsic</u> <u>core</u> of the convex set A, which is defined by $icr(A) := \{a \in A \mid \text{for each } x \in A/\{a\} \text{ there exists } y \in A \text{ such that } a \in (x, y)\}$, where (x, y) denotes the open line segment joining x and y. Namely, if $a \in icr(A)$ it is possible to move linearly from any point in A past a and remain in A. It coincides with the topological interior of A if Y is a finite dimensional Euclidean space and $A \subset Y$ is convex.

12/ If $\lim_{t\downarrow 0}\frac{F(a^\circ+ta) - F(a^\circ)}{t} = \lim_{t\uparrow 0}\frac{F(a^\circ+ta) - F(a^\circ)}{t}$ for all $a \in Y$, the linear functional Φ which maps any $a \in Y$ to this two-sided limit is called the gradient of F at a°, and denoted by $\nabla F(a^\circ)$.

13/ A proof of this lemma in finite-dimensional Euclidean spaces was given in Kobayashi [1983]. The current version using Pshenichnii's theorem was suggested by Darrell Duffie.

14/ Pshenichnii's theorem asserts the following: If Y is a locally convex, linear topological space, $A \subset Y$ is a convex set, f is a concave function on A, and f is continuous at a point $a^\circ \in A$, then $f(a^\circ) = \max\{f(a) \mid a \in A\}$ if and only if there exists a supergradient Φ of f at a° such that $\Phi(a^\circ) = \max\{\Phi(a) \mid a \in A\}$. For the proof see Holmes [1975] pp. 87-88.

15/ A linear functional Φ on Y is called a supergradient of F at $a^\circ \in A$ if $\Phi(a - a^\circ) \geq F(a) - F(a^\circ)$ holds for all $a \in A$.

16/ For the former assertion see Holmes, p. 29 and for the latter see ibid., p. 23.

17/ The market does not prohibit resales of the financial assets. The important assumption here is that he never considers these possibilities when he makes portfolio decisions. Hence the assumption does not mean the nonexistence of a "secondary market" for each asset.

REFERENCES

[1] Blackwell, D.: "Positive Dynamic Programming," Proc. 5th Berkeley Sym. Math. Stat. 1, U. of California Press, Berkeley, 1967.

[2] Holmes, Richard B.: Geometric Functional Analysis and its Applications, Springer-Verlag, New York Inc., 1975.

[3] Kobayashi, T.: "Expectations Equilibrium in an Intertemporal Capital Market and the Growth Asset Pricing Model," The Journal of Economics (in Japanese), The Society of Economics, U. of Tokyo, 49 (1983), 2-27.

[4] Kobayashi, T.: "A No-Speculation Theorem in Intertemporal Capital Markets and Some Applications," in print [1986].

[5] Kreps, David M.: "Decision Problems with Expected Utility Criteria, I: Upper and Lower Convergent Utility," Mathematics of Operations Research, 2 (1977), 45-53.

[6] Ross, Sheldon M.: "Dynamic Programming and Gambling Models," Advances in Appl. Probability 6 (1974), 593-606.

[7] Strauch, Ralph E.: "Negative Dynamic Programming," Ann. Math. Statist. 37 (1966), 871-890.

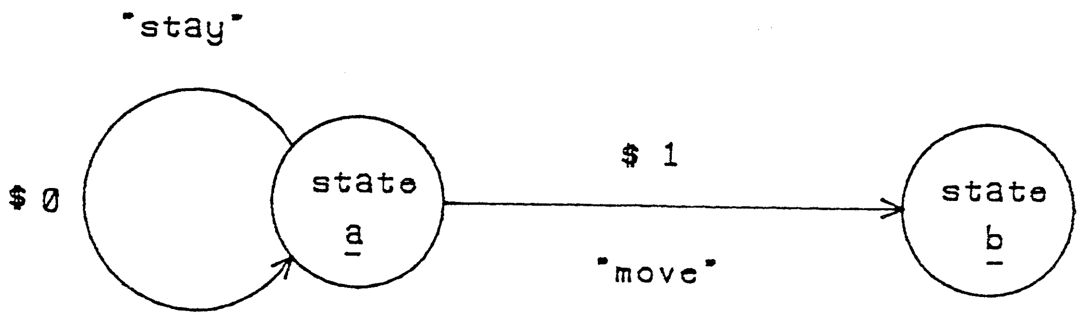


Fig. 1

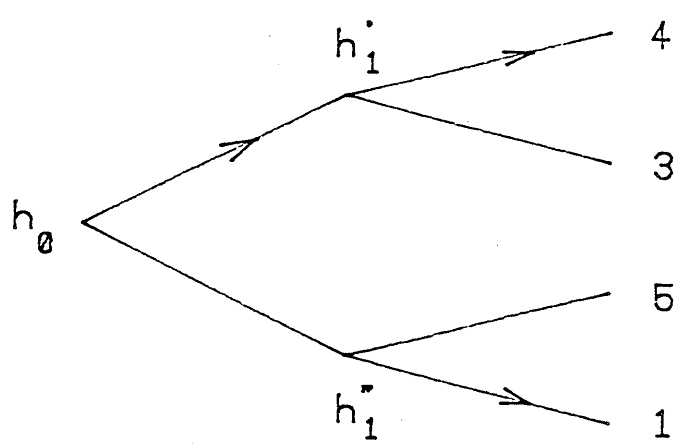
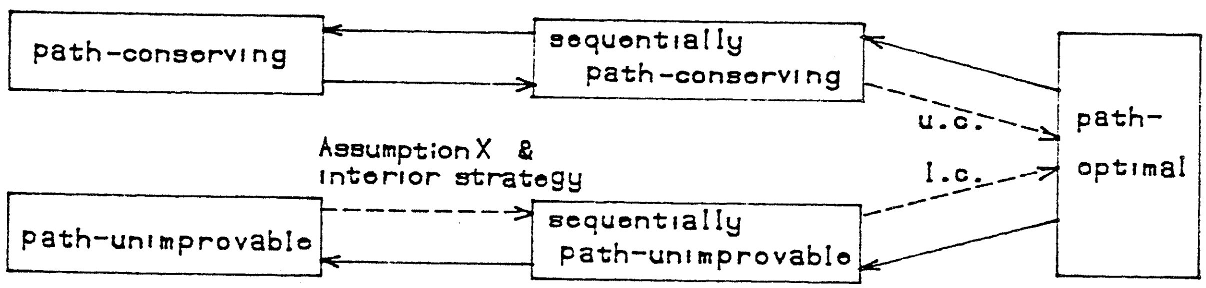


Fig. 2



path-conserving
sequentially path-conserving
path-unimprovable
sequentially path-unimprovable
Assumption X & interior strategy
u.c.
l.c.
path-optimal

Fig. 3