International Journal of Game Theory (1993) 22: 31–42

Bounded Variation of $\{V_n\}$ and its Limit[1]

EHUD LEHRER

Raymond and Beverly Sackler Faculty of Exact Sciences, School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel

Abstract: This note studies relations between $\lim V_n$ and the lower long-run average value ($\underline{V}$) in dynamic programming. It is shown that a certain bounded variation conditions of $\{V_n\}$ implies that $\lim V_n = \underline{V}$.

1 Introduction

This note deals with dynamic programming problems (DP). A decision maker observes the state of nature, s, and chooses an action, say a. In turn, he gets the immediate reward, $f(s, a)$, and the system switches to another state which depends stochastically on (s, a). Then, the decision maker observes the new prevailing state and chooses once again an action. He proceeds that way infinitely many times. A *policy* is a function from histories (which are finite strings of pairs of states and actions) to actions. Thus, any policy induces a probability distribution over infinite streams of rewards. Undiscounted DP and discounted or finite DP differ in the way infinite streams of rewards are evaluated. The evaluation of the stream is considered the payoff of the decision maker in the DP.

Undiscounted DP evaluate a stream of (expected) rewards by the limit of the finite averages. Usually the standard limit does not exist and, therefore, one must use either the upper limit (limsup) or the lower limit (liminf). The upper limit corresponds to the decision maker who considers primarily the highest averages he is about to experience, while the lower limit corresponds to one who cares most about the lowest averages about to occur. In a finite DP, only the average of a finite prefix of the sequence of rewards matters and all the remaining payments (in the tail of the sequence) are ignored.

A value of a DP at a fixed state, s, is defined as the highest achievable payoff in the respective DP, starting at the initial state s. The undiscounted values (lower and upper), the discounted and the one corresponding to the finite DP are denoted by $\underline{V}$, $\overline{V}$, V_λ (when the discount factor is λ) and V_n (when the length of the finite prefix considered is n), respectively. For general discussions about DP, the reader is referred to Blackwell (1962, 1965), and Blackwell, Freedman and Orkin (1974).

We study here the relation between $\underline{V}$ and $\lim V_n$. A similar problem was a subject of several papers in various configurations. Flynn (1974) studied the optimal

[1] Dov Monderer is acknowledged for helpful comments. I am also greatful to the two referees and the editor in charge of the *International Journal of Game Theory*.

0020-7276/93/1/31-42 $ 2.50

policies (under different measurements of effectiveness) rather than the resulting payoffs. Lehrer and Monderer (1989) dealt with the relation between $\overline{V}$ and limit of the discounted values. They proved that if the discounted value converges uniformly (on the state space), then this limit is equal to $\overline{V}$. It is also shown there that it does not in general coincide with the lower undiscounted value, $\underline{V}$. Lehrer and Sorin (1992) proved that a uniform convergence of V_n is equivalent to a uniform convergence of the discounted values. Moreover, the limits coincide.

The recent trend of papers is motivated by the seminal work of Mertens and Neyman (1981). They proved that a certain bounded variation (BV) condition on discounted values along admissible sequences of discount factors, λ_k, implies that the lower value of the undiscounted zero-sum stochastic game is equal to $\lim V_{\lambda_k}$. No similar result for a BV condition on V_n was given.

Here we provide a BV condition on V_n which implies that the lower value in infinite DP is equal to $\lim V_n$. We say that a sequence of integers $\{n_k\}$ is good if every element $\{n_k\}$ can be approximated by bounded summation of its ℓ predecessors. We show that if $\sum \| V_{n_k} - V_{n_{k+1}} \|$ is finite for $\{n_k\}$-good sequence then $\underline{V} = \lim_k V_{n_k}$.

The result of Mertens and Neyman (1981) obviously implies that if $\{V_{\lambda_k}\}$ satisfies BV for an admissible sequence V_{λ_k}, then $\underline{V} = \lim V_{\lambda_k}$. Had we known to show that BV of V_{n_k} on $\{n_k\}$-good implies BV of V_{λ_k} on admissible $\{\lambda_k\}$ we could have used both Mertens and Neyman (1981) and Lehrer and Sorin (1992) results to show our present result. The relations between BV of $\{V_n\}_n$ and BV $\{V_{\lambda_k}\}$, however, is yet to be investigated.

2 The Deterministic DP: Definitions and Notations

Let S be the set of states, and A be a set of actions. The reward function is $f: S \times A \to \mathbb{R}$. We assume that f is bounded. Without loss of generality f is bounded between 0 and 1. Finally, the deterministic transition function, τ, dictates the next prevailing state based on both: the current one and the action taken. Formally, the transition function, τ, is a function from $S \times A$ to S.

A deterministic dynamic programming problem (DDP) is defined by (S, s_0, A, f, τ), where $s_0 \in S$ is the initial state. A history of length t is an element in $S \times (S \times A)^t = H_t$. Denote $H = \bigcup_{t=0}^{\infty} H_t$. A policy, σ, is a function from H to A. σ specifies what should the decision maker do at any time and after every history. Since the transition function is deterministic, any policy, σ, induces a sequence of states $s_0, s_1, s_2, \ldots$ satisfying $\tau(s_0, \sigma(s_0)) = s_1$, $\tau(s_1, \sigma(s_0, s_1, \sigma(s_0)) = s_2$, and so forth. In other words, s_t is the prevailing state at time t if the policy σ is followed. We denote by $s_t(\sigma)$ the state s_t and by $a_t(\sigma)$ the action taken at time t. Thus, the immediate reward at time t is $x_t(\sigma) = f(s_t(\sigma), a_t(\sigma))$. Define, for every $s \in S$, the following value of the DDP with the initial state s

$$V_n(s) = \sup_{\sigma} (1/n) \sum_{t=1}^{n} x_{t-1}(\sigma),$$

where the supremum is taken over all the policies σ. V_n is the value of the finite DDP.

The lower long-run value of the DDP with the initial state s is defined as:

$$\underline{V}(s) = \sup_{\sigma} \liminf_{T} (1/T) \sum_{t=1}^{n} x_{t-1}(\sigma).$$

In the sequel V_n and $\underline{V}$ will be referred to as real functions defined on the set of states, S.

A play $h=(h_0, h_1, \ldots)$ at state s_0 is a sequence of pairs, $h_i=(s_i, a_i)\in S\times A$, where $s_{i+1}=\tau(s_i, a_i)$. Sometimes by denoting h_i we refer only to s_i. We say that s' *follows* s if there is a play h at s, m and a s.t. $h_m=(s', a)$.

Notice that in the case where s' follows s $kV_k(s')\leq kV_k(s)+m$ and therefore $V_k(s')-\frac{m}{k}\leq kV_k(s)$ for every integer k. It follows that if s' follows s, then $\limsup_k V_{n_k}(s')\leq \limsup_k \sup V_{n_k}(s)$ for every sequence $\{n_k\}$. This is a key property of dynamic programming problems and plays a central role in the proofs that follow.

Notation 1. Suppose that $h=(h_0, h_1, \ldots)$ is a play at s.

a. $f(h)$ denotes the sequence $\{f(h_i)\}_{i\in\mathbb{N}}$.
b. For any[2] $\ell, k\in\mathbb{N}$, where $\ell<k$ denote $A_{\ell,k}(f(h))=(f(h_\ell)+\ldots+f(h_{k-1}))/(k-\ell)$. For $\ell=0$ we omit the ℓ and denote it by $A_k(f(h))$.
c. For $t\in\mathbb{R}_+$, the play h^t denotes $(h_r, h_{r+1}, \ldots)$, where $r=\mathrm{Max}\{n<t \mid n\in\mathbb{N}\}+1$.
d. Denote $u_k=\| V_{n_k}-V_{n_{k+1}}\|_\infty$.

3 The Finitely Truncated DDP Values and $\underline{V}$

In this section a connection between the finitely truncated DDP values and the lower long-run value is demonstrated. Let $\{n_k\}_k\subseteq\mathbb{N}$. We say that the DDP satisfies[3] $\{n_k\}_k$-BV if

$$\sum_k \| V_{n_k}-V_{n_{k-1}}\|_\infty<\infty.$$

Denote $V_\infty=\lim_k V_{n_k}$.

Proposition 1. If DDP satisfies $\{n_k\}$-BV, then for every $\varepsilon>0$ there is M s.t. every $M\leq m$ satisfies

---

[2] here and in the sequel $\mathbb{N}$ denotes the set of non-negative integers.
[3] BV for *b*ounded *v*ariation.

$$V_m(s) \leq V_\infty(s) + \varepsilon \quad \textit{for every } s.$$

Proof: Otherwise there is a constant $c>0$, an increasing sequence m_ℓ of integers and a sequence s_ℓ of states s.t.

$$V_{m_\ell}(s_\ell) > V_\infty(s_\ell) + c.$$

For every ℓ take a play h_ℓ at s_ℓ s.t.

$$A_{m_\ell}(f(h_\ell)) \geq V_{m_\ell}(s_\ell) - c/2 > V_\infty(s_\ell) + c/2.$$

By a method described in the proof of Proposition 2 of (Lehrer & Sorin, 1992) one can find an integer, say J_ℓ, s.t.

$$A_{J_\ell, J_\ell + i}(f(h_\ell)) \geq V_{m_\ell}(s_\ell) - 3c/4 \quad \text{for every } 0 \leq i \leq \frac{cm_\ell}{4}.$$

If m_ℓ is sufficiently large than $\frac{c \cdot m_\ell}{4}$ is greater than n_k that satisfies $\| V_\infty - V_{n_k} \| < c/4$. If we let s'_ℓ be the n_k-th state in the play h_ℓ we get,

$$V_{n_k}(s'_\ell) \geq A_{J_\ell, J_\ell + n_k}(f(h_\ell)) \geq V_{m_\ell}(s_\ell) - 3c/4 \geq v_\infty(s_\ell) + c/4.$$

Thus,

$$V_\infty(s'_\ell) > V_{n_k}(s'_\ell) - \frac{c}{4} \geq V_\infty(s_\ell).$$

This is a contradiction since s'_ℓ follows s_ℓ. //

Let $\{n_k\}$ be an increasing sequence of integers. Denote $D_{k,\ell} = \{n_{k-1}, n_{k-2}, \ldots, n_{k-\ell}\}$. Denote,

$$D^m_{k,\ell} = \{x_1 + x_2 + \ldots + x_r \mid x_i \in D_{k,\ell}, r \leq m\}.$$

$D^m_{k,\ell}$ consists of summations of at most m numbers taken from the set $\{n_{k-1}, \ldots, n_{k-\ell}\}$.

Definition. A sequence $\{n_k\}$ is (m, ℓ)-*good* if it is increasing and if[4]

$$\sum_k \operatorname{dist}(n_k, D^m_{k,\ell})/n_k < \infty.$$

A sequence $\{n_k\}$ is good if it is (m, ℓ)-good for some $(m, \ell) \in \mathbb{N}^2$.

---

[4] $\operatorname{dist}(a, A) = \min_{b \in A} |a - b|$ for every $a \in \mathbb{N}$ and $A \subseteq \mathbb{N}$.

Examples.

1. The sequence $\{a^k\}_k$, where $a \in \mathbb{N}$ is $(a, 1)$-good because $D_{k,1} = \{a^{k-1}\}$ and $a^k \in D^a_{k,1} = \{a^k\}$. Thus, $\operatorname{dist}(a^k, D^a_{k,1}) = 0$ for every k.
2. The Fibonacci sequence $(n_k = n_{k-1} + n_{k-2})$ is $(2, 2)$-good because $D_{k,2} = \{n_{k-1}, n_{k-2}\}$ and $n_k = n_{k-1} + n_{k-2} \in D^2_{k,2}$ for every $k \geq 3$.

Remark 1. If $\{n_k\}$ is (m, ℓ)-good then $\limsup n_k/n_{k-1}$ is bounded by m.

Theorem 1. Suppose that $\{n_k\}$ is a good. If a DDP satisfies $\{n_k\}$-BV, then

$$\underline{V} = \lim_k V_{n_k} \left(\stackrel{\text{def}}{=} V_\infty\right).$$

Corollary 1. If a DDP satisfies $\{k\}_k$-BV, then $\underline{V} = \lim_k V_k$.

Proof: If DDP satisfies $\{m_k\}$-BV and $\{n_k\}$ is a subsequence of $\{m_k\}$, then DDP also satisfies $\{n_k\}$-BV. If, in addition, $\{n_k\}$ is good, then by Theorem 1 $\underline{V} = \lim V_{n_k} = V_\infty$. In our case every $\{n_k\}$-good sequence is a subsequence of $\{k\}_k$. //

Corollary 2. If a DDP satisfies $\{n_k\}$-BV and $\{n_k\}$ is good, then $\lim V_n$ *exists and it is equal to $\underline{V}$.*

Proof: Clearly $\underline{V} \leq \liminf_n V_n$. By Proposition 1 $V_\infty \geq \limsup_n V_n$. Since $\underline{V} = V_\infty$ the proof is complete. //

4 An Outline of the Proof of Theorem 1

Since the proof involves a construction which uses a lot of parameters, I will present a proof for a particular $(2, 2)$-good sequence: a Fibonacci sequence where $n_{k+1} = n_k + n_{k-1}$. Using this particular sequence, the proof becomes more transparent. For the general proof, one can use exactly the same ideas. The necessary modification of the proof will be indicated at the end.

The fact that $\underline{V} \leq V_\infty$ is obvious. Thus, it is enough to show that for every $\varepsilon > 0$ there is a play h at s_0 which satisfies $\liminf A_t(h) \geq V_\infty(s_0) - \varepsilon$. The idea is to construct a play h inductively. At the k-th step of the inductive procedure, a finite sequence of n_k consecutive states will be added. The average of the rewards over these n_k states is high enough. Moreover, any intermediate average (of less than n_k rewards), however low, cannot affect the total average by much, because of two reasons. First, the payoffs are bounded and, second, the ratio between n_k and the numbers of its predecessors $(\sum_{\ell < k} n_\ell)$ is uniformly bounded.

The method of finding at the k-th step the n_k states is the following. Suppose that at the end of the former step the state s was reached. We first take a play $h(k)$ at s, which almost (up to δ_k) realizes $V_{n_{k+2}}(s)$. I.e.,

$$V_{n_{k+2}}(s)-\delta_k \leq A_{n_{k+2}}(f(h(k))).$$

The average of the n_{k+2} first rewards along the play $h(k)$, $A_{n_{k+2}}(f(h(k)))$, can be written as a combination of two averages. The first is the average of the first n_k rewards and the second is the average of the last n_{k+1} rewards. The goal of the proof is two-fold: (i) To show that the first average is high enough. Therefore, the corresponding n_k states will be annexed as a part of the play h. (ii) To show that the second average is sufficiently high. The latter will ensure that at the (n_k+1)-th state of the play $h(k)$, say, s', the value $V_{n_{k+1}}$ is high enough and therefore (because of the BV condition), $V_{n_{k+3}}(s')$ is also relatively high. This enables one to proceed with the next step of the inductive process (with the state s' and $V_{n_{k+3}}$).

Notice that $n_{k+1}/n_k \leq 2$, and therefore the weight of the second average is at least half the weight of first one. The first average is not smaller than $V_{n_{k+2}}(s)-\varepsilon$ because, if the opposite happens, then the second average (and, therefore, $V_{n_{k+1}}(s')$) is at least $V_{n_{k+2}}(s)+\varepsilon/2$. This will lead to a contradiction by the following argument. Suppose that $V_{n_{k+2}}(s)$ was proved (as a part of the inductive process) to be strictly greater than $V_\infty(s_0)-\varepsilon/4$, where s_0 is the initial state. Thus, $V_{n_{k+1}}(s')$ is strictly greater than $V_\infty(s_0)+\varepsilon/4$. Suppose, furthermore, that the series $\sum_{\ell=k}^{\infty} u_\ell$ is smaller than $\varepsilon/4$. Hence, $V_{n_{k+1}}(s')$ is at most $V_\infty(s')+\varepsilon/4$. Thus, $V_\infty(s')>V_\infty(s_0)$, which is impossible since s' follows s_0. This argument proves (i).

To establish (ii) notice that $V_{n_k}(s)$ is close to $V_{n_{k+2}}(s)$ (because of BV). Thus, the first average ($A_{n_k}(f(h))$ is not too high) and, therefore, the second average is not too low.

An important point should be noticed here. When we pass from one step to another, we may lose height. (In the above notation $V_{n_{k+2}}(s)$ may be smaller than $V_{n_k}(s)$.) The loss may accumulate to more than ε at the limit. The BV condition, however, rules out this possibility.

5 Proof of Theorem 1

Let $\varepsilon>0$. W.l.o.g., $\sum u_k<\varepsilon/8$. Take $\{\delta_k\}$, a sequence of positive numbers satisfying $\sum \delta_k<\varepsilon/8$. We assume that $V_\infty(s_0)-\varepsilon/2 \leq V_{n_3}(s_0)$ and that $h(1)=(h_0(1), h_2(1), \ldots, h_i(1), \ldots)$ is a play at s_0 which satisfies

$$V_\infty(s_0)-\varepsilon/2-\delta_1 \leq V_{n_3}(h_0(1))-\delta_1 \leq A_{n_3}(f(h(1))). \tag{1}$$

(Formally $h_0(1)$ is a pair: a state and an action. Here and in the sequel we ignore the second component and we refer only to the state.) $A_{n_3}(f(h(1)))$ can be written as a combination of two averages with length of n_1 and n_2:

$$A_{n_3}(f(h(1))) = (n_1/n_3)A_{n_1}(f(h(1))) + (n_2/n_3)A_{n_1,n_3}(f(h(1))). \tag{2}$$

$$A_{n_1}(f(h(1))) \le V_{n_1}(h_0(1)) \le V_{n_3}(h_0(1)) + u_1 + u_2. \tag{3}$$

Thus, from (1)–(3) one obtains

$$V_{n_3}(h_0(1)) - \delta_1 \le (n_1/n_3)(V_{n_3}(h_0(1)) + u_1 + u_2) + (n_2/n_3)A_{n_1,n_3}(f(h(1))). \tag{4}$$

Rearranging (4) we get

$$V_{n_3}(h_0(1)) - (n_3/n_2)\delta_1 - (n_1/n_2)(u_1+u_2) \le A_{n_1,n_3}(f(h(1))) \le V_{n_2}(h_{n_1}(1)).$$

Therefore,

$$V_{n_3}(h_0(1)) - 2\delta_1 - (u_1+u_2) \le V_{n_2}(h_{n_1}(1)). \tag{5}$$

The Second Step of the Induction: Let $h(2)$ be a play at $h_{n_1}(1)$ (in particular, $h_0(2) = h_{n_1}(1)$) satisfying

$$V_{n_4}(h_0(2)) - \delta_2 \le A_{n_4}(f(h(2))).$$

By a similar derivation to the above (see (5)), one gets

$$V_{n_4}(h_0(2)) - 2\delta_2 - (u_2+u_3) \le V_{n_3}(h_{n_2}(2)). \tag{6}$$

The k-th step of the Induction: At this step we take a play $h(k)$ at $h_{n_{k-1}}(k-1)$ satisfying

$$V_{n_k+2}(h_0(k)) - \delta_k \le A_{n_k+2}(f(h(k))) \tag{7}$$

using the fact that $\| V_{n_k} - V_{n_k+2} \| \le u_k + u_{k+1}$, one gets (compare with (6))

$$\begin{aligned} V_{n_k}(h_0(k)) &- 2\delta_k - 2(u_k+u_{k+1}) \\ &\le V_{n_k+2}(h_0(k)) - 2\delta_k - (u_k+u_{k+1}) \le V_{n_{k+1}}(h_{n_k}(k)). \end{aligned} \tag{8}$$

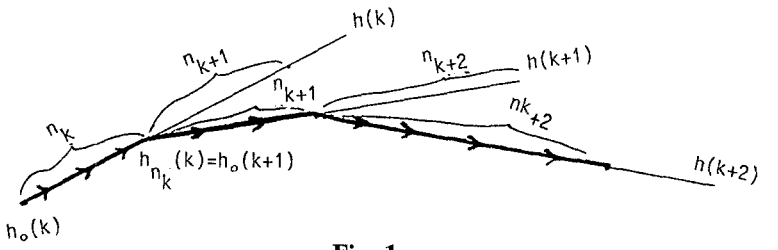

Fig. 1

Our goal is to prove that the lower limit of the finite averages of the sequence $b \stackrel{\text{def}}{=} (f(h_0(1)), f(h_1(1)), \ldots, f(h_{n_1-1}(1), f(h_0(2)), f(h_1(2)), \ldots, f(h_{n_2-1}(2)), f(h_0(3)), f(h_1(3)), \ldots)$ is at least $V_\infty(s_0) - 4\varepsilon$.

The proof is carried out through six lemmas.

Lemma 1. For every k

$$V_{n_3}(s_0) - 2\sum_{\ell=1}^{k} \delta_\ell - 2\sum_{\ell=1}^{k} (u_\ell + u_{\ell+1}) \leq V_{n_{k+1}}(h_{n_k}(k)).$$

Proof: By induction, using (8). //

Lemma 2. $V_\infty(s_0) - \varepsilon \leq V_{n_{k+1}}(h_{n_k}(k))$ *for all k.*

Proof: By Lemma 1 and the choice of parameters. //

Lemma 3. $V_{n_{k+2}}(h_0(k)) - \varepsilon \leq A_{n_k}(f(h(k)))$.

Proof: Suppose that the lemma does not hold. In this case,

$$\begin{aligned} A_{n_{k+2}}(f(h(k))) &= (n_{k+1}/n_{k+2}) A_{n_k, n_{k+2}}(f(h(k))) + (n_k/n_{k+2}) A_{n_k}(f(h(k))) \\ &< (n_{k+1}/n_{k+2}) A_{n_k, n_{k+2}}(f(h(k))) + (n_k/n_{k+2})(V_{n_{k+2}}(h_0(k)) - \varepsilon). \end{aligned}$$

Therefore (by (7)),

$$\begin{aligned} &(n_{k+2}/n_{k+1})[V_{n_{k+2}}(h_0(k)) - \delta_k - (n_k/n_{k+2})(V_{n_{k+2}}(h_0(k)) - \varepsilon)] \\ &\qquad \leq A_{n_k, n_{k+2}}(f(h(k))) \stackrel{\text{def}}{=} A. \end{aligned}$$

Hence,

$$V_{n_{k+2}}(h_0(k))[n_{k+2}/n_{k+1} - n_k/n_{k+1}] - 2\delta_k + \varepsilon(n_k/n_{k+1}) \leq A.$$

Thus,

$$\begin{aligned} V_{n_{k+2}}(h_0(k)) - 2\delta_k + \varepsilon(n_k/n_{k+1}) &\leq A \leq V_{n_{k+1}}(h_{n_k}(k)) \\ &\leq V_{n_{k+2}}(h_{n_k}(k)) + u_{k+1} \leq V_\infty(h_{n_k}(k)) + \sum_{k<\ell} u_\ell. \end{aligned}$$

We conclude that

$$V_{n_{k+2}}(h_0(k)) - 2\delta_k + \varepsilon(n_k/n_{k+1}) - \varepsilon/8 \leq V_\infty(h_{n_k}(k)).$$

Since $V_\infty(h_0(k)) - \varepsilon/8 \leq V_{n_{k+2}}(h_0(k))$, and since w.l.o.g. $n_k/n_{k+1} > \frac{1}{2}$ we obtain $V_\infty(h_0(k)) < V_\infty(h_{n_k}(k))$. This is impossible because $h_{n_k}(k)$ follows $h_0(k)$. //

Before we proceed to the next lemma, recall that b is the sequence of payoffs along the infinite play path we have previously constructed.

Lemma 4. Denote $t_k = n_1 + n_2 + \ldots + n_k$. The finite averages of length t_k of the sequence b are at least $V_\infty(s_0) - 3\varepsilon$.

Proof: By the previous lemma, the t_k-period averages are at least $\inf_k \{V_{n_{k+2}}(h_0(k)) - \varepsilon\}$. However, for all k,

$$\begin{aligned} V_{n_{k+2}}(h_0(k)) - \varepsilon &\geq V_\infty(h_0(k)) - \varepsilon - \sum_{\ell \geq k+2} u_\ell \geq V_\infty(h_{n_k}(k)) \\ &\quad - \varepsilon - \varepsilon/8 \geq V_{n_{k+1}}(h_{n_k}(k)) - \varepsilon - \varepsilon/8 - \varepsilon/8 \\ &= V_{n_{k+1}}(h_{n_k}(k)) - (1.25)\varepsilon \qquad \text{(by Lemma 2)} \\ &\geq V_\infty(s_0) - 3\varepsilon. \quad // \end{aligned}$$

For the next lemma recall the definition of $h(k)$ in (7).

Lemma 5. For every $1/2 > \eta' > \eta > 0$, if $\eta n_{k+2} < \ell < (1-\eta') n_{k+2}$, then $A_\ell(f(h(k))) \geq V_{n_{k+2}}(h_0(k)) - \alpha_k$, where $\alpha_k \to 0$ as $k \to \infty$.

Proof: Otherwise, for infinitely many k's and ℓ's ($\eta n_{k+2} < \ell < (1-\eta') n_{k+2}$) the opposite inequality holds for a fixed $c > 0$. Namely,

$$A_\ell(f(h(k))) < V_{n_{k+2}}(h_0(k)) - c \tag{9}$$

for infinitely many k's. Therefore,

$$V_{n_{k+2}}(h_0(k)) - \delta_k \leq A_{n_{k+2}}(f(h(k))) \leq (1-\eta)A + \eta(V_{n_{k+2}}(h_0(k)) - c),$$

where $A = A_{\ell, n_{k+2}}(f(h(k)))$.

Thus,

$$V_{n_{k+2}}(h_0(k)) - \delta_k/(1-\eta) + \eta c/(1-\eta) \leq A.$$

As $\delta_k \to 0$, $V_{n_{k+2}}(h_0(k)) + \eta c \leq A$ for k large enough. Since $n_{k+2} \to \infty$, by Proposition 1, there is k s.t. whenever $m \geq \eta' n_{k+2}$ the following holds: $V_m - V_\infty < \eta c/2$. Therefore,

$$V_\infty(h_0(k)) + \eta c/2 < A \leq V_{n_{k+2}-\ell}(h_\ell(k)) \leq V_\infty(h_\ell(k)) + \eta c/2.$$

(The last inequality holds because $n_{k+2} - \ell > \eta' n_{k+2}$.) Hence, $V_\infty(h_0(k)) < V_\infty(h_\ell(k))$, which is a contradiction since the state $h_\ell(k)$ follows the state $h_0(k)$. //

Lemma 6. $\liminf_t A_t(b) \geq V_\infty(s_0) - 4\varepsilon$.

Proof: By Lemma 4, the t_k-partial averages are greater than $V_\infty(s_0) - 3\varepsilon$. For $t_k < t < t_{k+1}$

$$A_t(b) = (t_k/t) A_{t_k}(b) + ((t - t_k)/t) A_{t_k, t}(b)$$

$$\geq (t_k/t)(V_\infty(s_0)-3\varepsilon)+((t-t_k)/t)A_{t_k,t}(b).$$

Three Cases.

First: $(t-t_k)/n_{k+1}\leq\varepsilon$. Thus, $(t-t_k)/t\leq\varepsilon$ and $t_k/t\geq 1-\varepsilon$. Hence,

$$A_t(b)\geq(1-\varepsilon)(V_\infty(s_0)-3\varepsilon)\geq V_\infty(s_0)-4\varepsilon.$$

Second: $\varepsilon(n_{k+1})\leq t-t_k\leq(1-\varepsilon)n_{k+1}$. Thus $\left(\frac{n_{k+1}}{n_{k+3}}\right)\varepsilon n_{k+3}\leq t-t_k\leq(1-\varepsilon)n_{k+1}\leq$ $(1-\varepsilon)n_{k+3}$. By using Lemma 5 with $\eta'=\varepsilon$ and $\eta=\dfrac{n_{k+1}}{n_{k+3}}\varepsilon$ one obtains

$$A_{t_k,t}(b)=A_{t-t_k}(f(h(k+1)))\geq V_{n_{k+3}}(h_0(k+1))-\alpha_{k+1}.$$

Therefore

$$A_t(b)\geq(t_k/t)(V_\infty(s_0)-3\varepsilon)+((t-t_k)/t)(V_{n_{k+3}}(h_0(k+1))-\alpha_{k+1}). \tag{10}$$

Now, by the proof of Lemma 4,

$$V_{n_{k+3}}(h_0(k+1))\geq V_\infty(s_0)-2\varepsilon.$$

Hence, by (10), we derive

$$A_t(b)\geq(t_k/t)(V_\infty(s_0)-3\varepsilon)+((t-t_k)/t)(V_\infty(s_0)-2\varepsilon-\alpha_{k+1})\geq V_\infty(s_0)-3\varepsilon,$$

when α_{k+1} is small enough.

Third: $t-t_k\geq(1-\varepsilon)n_{k+1}$. Now we divide $A_t(b)$ into two partial averages (from 0 to $t_k+[(1-\varepsilon)n_{k+1}]$ and from $t_k+[(1-\varepsilon)n_{k+1}]+1$ to t. The first average which has a weight of at least $1-\varepsilon$ is, by the former argument, at least $V_\infty(s_0)-3\varepsilon$. Therefore, $A_t(b)\geq(1-\varepsilon)(V_\infty-3\varepsilon)\geq V_\infty-4\varepsilon$. //

In the remainder of this section we indicate the modifications needed to prove the general case. In case $\{n_k\}$ is (m,ℓ)-good then any n_k can be approximated by a summation $r_k=\sum_{i=1}^{m_k} r_k^i$, where $m_k\leq m$ and $r_k^i\in\{n_{k-1},\ldots,n_{k-\ell}\}$ for any i. Denote $d_k=|n_k-r_k|/n_k$. By assumption $\Sigma d_k<\infty$. We may assume that Σd_k is as small as needed. We start the inductive process from time $K\geq m$ that will be determined later. Suppose that state s is reached at time $k-1$. At time k a play $h(k)$ at s is found so that

$$A_{n_k}(f(h(k)))\geq V_{n_k}(s)-\delta_k.$$

By Remark 1 there is $0<c$ s.t. $\frac{n_{k-\ell}}{n_k} > c$. Since $d_k \to 0$ we may assume that $\frac{n_{k-\ell}}{r_k}$ is also greater than c. The average $A_{n_k}(f(h(k)))$ is divided into the partial averages corresponding to r_k^i $(i=1, \ldots, m_k)$. By a similar argument to the one in (2), (3), (4) one may deduce that each one of these averages, in particular, the second one (of length r_k^2) is at least $A_{n_k}(f(h(k))) - \delta_k - C(d_k + u_{k-1} + \ldots + u_{k-\ell})$ with some constant C. This means that

$$V_{r_k^2}(h_{r_k^1}(k)) \geq V_{n_k}(h_0(k)) - \delta_k - C'(d_k + u_{k-1} + \ldots + u_{k-\ell}),$$

with some constant C'. But the left side is at most $V_{n_{k+1}}(h_{r_k^1}(k)) + u_k + \ldots + u_{k-\ell}$. Therefore we get (compare with (8))

$$V_{n_{k+1}}(h_{r_k^1}(k)) \geq V_{n_k}(h_0(k)) - \delta_k - C''(d_k + u_k + \ldots + u_{k-\ell}) \quad \forall k \geq k, \tag{11}$$

where C'' is a constant. Now the inductive process proceeds with the state $h_{r_k^1}(k)$ (this is the state reached at time k): a play $h(k+1)$ at $h_{r_k^1}(k)$ is found and so forth.

Lemma 2 holds because the induction employed in the proof of Lemma 1 can be applied for (11), and because the series $\sum_{k \geq K} \delta_k + C''(d_k + u_k + \ldots + u_{k-\ell})$ converges (so w.l.o.g. it can be assumed to be as close as needed to 0). The proof of Lemma 3 relies on the fact that $\frac{n_k}{n_{k+1}}$ is uniformly bounded from below. The same argument can be proved here since $\frac{n_{k-\ell}}{n_k}$ is uniformly bounded from below. Lemmas 4 and 5 hold without any change. Finally Lemma 6 relies once again on the boundedness of $\frac{n_{k+1}}{n_k}$ and it holds here as well.

Conjecture. If a DDP satisfies $\{n_k\}$-BV and $\sup(n_{k+1}/n_k) < \infty$, then $\underline{V} = \lim_k V_{n_k}$.

6 The Stochastic DP

The former results can be generalized to stochastic DP in the following manner. Consider a countable-Borel model in which S is countable state space, A is a Borel set of actions and q is a transition probability from $S \times A$ to S. Finally, f is a measurable bounded payoff function defined on $S \times A$.

Every Markov strategy $\sigma = (\sigma_1, \sigma_2, \ldots)$ (Blackwell, 1965) induces a sequence, $\{w_n\}$, of probability distribution over S as follows. w_0 is the Dirac measure on $s_0 a_0$ $w_{n+1}(S') = \int_S q(S' \mid s, \sigma_n(s))\, w_n(dt)$ for every Borel set $S' \subseteq S$. The corresponding payoff sequence, $\{x_n(\sigma)\}$ is defined as

$$x_n(\sigma) = \int_S f(s, \sigma_n(s))\, w_n(dt).$$

Define now $\underline{V}^M$, the lower long-run average value as $\inf_\sigma E[(1/T)\sum_{t=1}^{T} x_n(\sigma)]$, where the infimum is taken over all Markov strategy. The finite values, V_n, are defined in the usual way. Moreover, the function V_n, defined on S, can be extended to all the probability distribution over S.

It can be easily verified that if (BV) holds for $\{V_n\}$ when defined on S, then (BV) holds also when $\{V_n\}$ are extended to the larger domain of probability distributions over S. The Theorem and the proof method given above work for countable-Borel model with the sole change that $\underline{V}^M$ replaces $\underline{V}$.

For non-countable state spaces, either an assumption of the existence of an ε-optimal strategy for every n-truncated problem (Lehrer and Monderer, 1990) or a measurable selection theorem are needed. In the second case the set-up of (Blackwell, Freedman and Orkin, 1974) can be used.

References

Blackwell D (1962) "Discrete Dynamic Programming", Annals of Mathematical Statistics 33, 719–726.

Blackwell D (1965) "Discounted Dynamic Programming", Annals of Mathematical Statistics 36, 226–235.

Blackwell D, Freedman D and Orkin M (1974) "The Optimal Reward Operator in Dynamic Programming", Annals of Probability 2, 926–941.

Flynn J (1974) "Averaging Vs. Discounting in Dynamic Programming: A Counter Example", The Annals of Statistics 2, 411–413.

Lehrer E and Monderer D (1989) "Discounting Vs. Averaging in Dynamic Programming", to appear in Games and Economic Behavior.

Lehrer E and Monderer D (1990) "Low Discounting and The Upper Long-Run Average Value in Dynamic Programming", to appear in Games and Economic Behavior.

Lehrer E and Sorin S (1992) "A Uniform Tauberian Theorem in Dynamic Programming", Mathematics of Operations Research 17, 2, 303–307.

Mertens J-F and Neyman A (1981) "Stochastic Games", International Journal of Game Theory 10, 2, 53–66.

Received April 1992
Revised version February 1993