MATHEMATICS OF OPERATIONS RESEARCH
Vol. 38, No. 1, February 2013, pp. 108–121
ISSN 0364-765X (print) | ISSN 1526-5471 (online)
http://dx.doi.org/10.1287/moor.1120.0561

Persistently Optimal Policies in Stochastic Dynamic Programming with Generalized Discounting

A. Jaśkiewicz
Institute of Mathematics and Computer Science, Wrocław University of Technology,
50-370 Wrocław, Poland, anna.jaskiewicz@pwr.wroc.pl

J. Matkowski, A. S. Nowak
Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra,
65-246 Zielona Góra, Poland {j.matkowski@wmie.uz.zgora.pl, a.nowak@wmie.uz.zgora.pl}

In this paper we study a Markov decision process with a nonlinear discount function. First, we define a utility on the space of trajectories of the process in the finite and infinite time horizon and then take their expected values. It turns out that the associated optimization problem leads to a nonstationary dynamic programming and an infinite system of Bellman equations, which result in obtaining persistently optimal policies. Our theory is enriched by examples.

Key words: stochastic dynamic programming; persistently optimal policies; variable discounting; Bellman equation; resource extraction; growth theory
MSC2000 subject classification: Primary: 90C40; secondary: 90C39, 91B32, 91B62
OR/MS subject classification: Primary: dynamic programming/optimal control: Markov infinite states; secondary: utility/preference: applications; decision analysis: applications
History: Received July 14, 2011; revised May 7, 2012. Published online in *Articles in Advance* October 24, 2012.

1. Introduction. Most of the literature on dynamic models in economics and operations research proceeds on the assumption that preferences are represented by a functional which is additive over time and discounts future rewards at a constant rate. More precisely, if $c := \{c_t\}$ is a feasible consumption sequence in an optimal growth model and the decision maker is equipped with a bounded instantaneous utility function u, then the standard *time-additive and separable utility* is given by

$$U(c) := \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{1}$$

where $\beta \in (0, 1)$ is a fixed discount factor; see Samuelson [33]. Time-additive separable utility has dominated research owing to the mathematical simplification derived from that functional form. However, this assumption regarding a constant discount coefficient is quite restrictive.

Koopmans [24] proposed a general approach to construction of recursive utility functions via the so-called *aggregator*. Such an aggregator, roughly speaking, is a function $G(a, r)$ of two real variables. A recursive utility U^* is then a solution to the following equation:

$$U^*(c_t, c_{t+1}, \ldots) = G(u(c_t), U^*(c_{t+1}, c_{t+2}, \ldots)), \tag{2}$$

for any t and $c = \{c_\tau\}$. This equation indicates weak separability between present and future consumption. Namely, a utility enjoyed from period t on depends on current consumption c_t and the aggregate utility $U^*(c_{t+1}, c_{t+2}, \ldots)$ from period $t+1$ onwards. Notably, U in (1) can be written as in (2) with the aid of aggregator

$$G(a, r) = a + \beta r. \tag{3}$$

In the literature, Equation (2) is referred to as Koopmans' equation and can be obtained by applying the Banach contraction principle. This is because one of the commonly used conditions imposed on G is:

$$|G(a, r_1) - G(a, r_1)| \le \beta |r_1 - r_2|,$$

for each a, r_1, r_2 and $\beta \in (0, 1)$; see Denardo [11], Lucas and Stokey [26], Stokey et al. [37].[1] Hence, Denardo [11], Lucas and Stokey [26] have shown that if G satisfies the above contraction assumption, then there is a unique bounded function U^* satisfying (2). Recursive utilities derived in this way need not possess time additivity and separability properties; see Becker and Boyd III [1], Boyd III [7].

[1] A similar contraction assumption for stochastic dynamic programming was used by Bertsekas [3] and Porteus [31].

In this paper we propose to replace a linear function $r \mapsto \beta r$ in (3) with an increasing real-valued function δ. More precisely, we shall consider a simple aggregator $G(a, r) = a + \delta(r)$, where δ satisfies some reasonable conditions. An extension of the Banach contraction principle given by Matkowski [27] allows us to obtain a utility of the form

$$U_\delta^*(c) = u(c_0) + \delta(u(c_1) + \delta(u(c_2) + \cdots)), \qquad c = \{c_t\}. \tag{4}$$

In this way, we construct a much larger class of recursive utilities. Clearly, (4) is not in general separable and reduces to (1), when $\delta(r) = \beta r$ for all r. But, if δ is merely piecewise linear, that is, for some constants $\beta_1, \beta_2 \in (0, 1)$, $\delta(r) = \beta_1 r$ for $r \geq 0$, and $\delta(r) = \beta_2 r$ for $r < 0$, then we deal with a nonseparable case. It is worth emphasizing that the discount functions δ may possess different features such as "magnitude effect" or "sign effect." The reader interested in these particular properties is referred to Jaśkiewicz et al. [22], where a deterministic case was investigated.

Our paper is devoted to a study of recursive utilities within a stochastic framework. More precisely, we deal with the recursive utilities on the sample path space of a stochastic decision process with a Markovian transition law and then use their expectations with respect to probability measures (on the space of infinite histories) induced by policies. The model, that we are concerned with resembles nonstationary dynamic programming with a general utility function. Unfortunately, in such a setup we cannot expect to obtain a stationary or Markov optimal policy. However, we are able to prove the existence of *persistently optimal policies* and give their characterization by a system of Bellman's optimality equations. Let us mention that an optimal policy is called persistently optimal, if it is conditionally optimal after every finite history. Comparable problems and related results, but without variable discounting, were studied by Dubins and Savage [12], Dubins and Sudderth [13] (in gambling theory), Hinderer [20], Kertz and Nachman [23], Schäl [35, 36], Feinberg [15, 16] (in stochastic dynamic programming), and by Nowak [30] (in dynamic games). On one hand, our results for linear functions δ reduce to those of Bertsekas [3], Bertsekas and Shreve [4], Blackwell [6], Schäl [34], Strauch [38] (in discounted dynamic programming), Brock and Mirman [8], Stokey et al. [37] (in the theory of optimal economic growth) and, on the other hand, they generalize the papers of Strauch [38] and Schäl [34].[2] Indeed, our optimality results are obtained for dynamic models with a bounded-from-above instantaneous (one-period) utility function u. To solve such models we utilize a truncation technique. Namely, first we study models with the utility functions $u^s(c) := \max\{u(c), -s\}$, $s \geq 1$, and then letting s go to infinity, we obtain the existence of a persistently optimal policy in the original model with the one-period utility u. Basic convergence results to our derivation of the Bellman equations are given in §6 and can be read independently. Finally, we would like to point out that our approach is different from the one that uses "nonexpected utilities," e.g., compared with Kreps and Porteus [25], Becker and Boyd III [1] or Weil [39].

2. Preliminaries. Let N (R) denote the set of all positive integers (all real numbers) and let $N_0 = N \cup \{0\}$, $\underline{R} = R \cup \{-\infty\}$. A *Borel space* is a nonempty Borel subset of a complete separable metric space. Let X, A be Borel spaces. Assume that D is a Borel subset of $X \times A$ such that

$$A(x) := \{c \in A\colon (x, c) \in D\}$$

is nonempty for any $x \in X$. Then, it is well known that there exists a universally measurable mapping $\varphi\colon X \mapsto A$ such that $\varphi(x) \in A(x)$ for each $x \in X$; see the Yankov-von Neumann theorem in Bertsekas and Shreve [4] or Dynkin and Yushkevich [14]. If every set $A(x)$ is compact, then by Corollary 1 in Brown and Purves [9], the mapping φ may be Borel measurable. A set-valued mapping $x \mapsto A(x)$ (induced by the set D) is called *upper semicontinuous* if $\{x \in X\colon A(x) \cap K \neq \varnothing\}$ is closed for each closed set $K \subset A$.

LEMMA 1. (a) *Let $g\colon D \mapsto \underline{R}$ be a Borel measurable function such that $c \mapsto g(x, c)$ is upper semicontinuous on $A(x)$ for each $x \in X$. Then,*

$$g^*(x) := \max_{c \in A(x)} g(x, c)$$

is Borel measurable and there exists a Borel measurable mapping $f^*\colon X \mapsto A$ such that

$$f^*(x) \in \arg\max_{c \in A(x)} g(x, c)$$

for all $x \in X$.

[2] An excellent survey of various criteria in dynamic programming can be found in Feinberg [17] and Feinberg and Shwartz [18].

(b) *If, in addition, we assume that $x \mapsto A(x)$ is upper semicontinuous and g is upper semicontinuous on D, then g^* is also upper semicontinuous.*

Proof. Part (a) follows from Corollary 1 in Brown and Purves [9]. Part (b) is a corollary to Berge's maximum theorem; see pages 115–116 in Berge [2] and Proposition 10.2 in Schäl [34]. □

Let Y be a metric space. By $\hat{B}(Y)$ we denote the space of all bounded-from-above $\underline{R}$-valued Borel measurable functions on Y. The subspace of all bounded functions in $\hat{B}(Y)$ is denoted by $B(Y)$.

3. The dynamic programming model. A *discrete-time decision process* is specified by the following objects:

(i) X is the *state space* and is assumed to be a Borel space.
(ii) A is the *action space* and is assumed to be a Borel space.
(iii) D is a nonempty Borel subset of $X \times A$. We assume that for each $x \in X$, the nonempty x-section

$$A(x) := \{c \in A\colon (x, c) \in D\}$$

of D represents the set of *actions available in state* x. (In the context of growth theory, $c \in A(x)$ is often referred to as a *feasible consumption* when the stock is $x \in X$.)

(iv) q is a *transition probability* from D to X.

Let $\hat{H} = (X \times A) \times (X \times A) \times \cdots$, and

$$\hat{H}_0 = X, \qquad \hat{H}_n = \underbrace{(X \times A) \times \cdots \times (X \times A)}_{n} \times X, \quad n \in N.$$

We assume that $\hat{H}_n$ and $\hat{H}$ are equipped with the product Borel σ-algebras. Let $D_{m+1} := D \times \cdots \times D$ ($m+1$ times) for $m \in N_0$. Define $H_0 = X$ and $H_n := D_n \times X$ for $n \in N$. Then H_n and H are the sets of all *feasible histories* $h_n = (x_0, c_0, \ldots, x_n)$ and $h = (x_0, c_0, x_1, c_1, \ldots)$, respectively, where $c_k \in A(x_k)$, $x_k \in X$ for each $k \in N_0$. Clearly, H_n (H) is a Borel subset of $\hat{H}_n$ ($\hat{H}$).

(v) $u \in \hat{B}(D)$ is a *one-period utility function.*
(vi) $\delta\colon \underline{R} \mapsto \underline{R}$ is a *discount function.*

A *policy* $\pi = \{\pi_n\}$ is defined as a sequence of universally measurable transition probabilities π_n from H_n to A such that $\pi_n(A(x_n) \mid h_n) = 1$ for each $h_n \in H_n, n \in N_0$. We write Π to denote the *set of all policies*. A *nonrandomized policy* $\pi = \{\pi_n\}$ is a sequence of universally measurable mappings $\pi_n\colon H_n \mapsto A$ such that $\pi_n(h_n) \in A(x_n)$ for each $h_n \in H_n, n \in N_0$. By Π_0 we denote the *set of all nonrandomized policies*. Clearly, Π_0 can be viewed as a subset of Π. By the Yankov-von Neumann theorem, the set Π_0 is nonempty; see Bertsekas and Shreve [4] or Dynkin and Yushkevich [14]. If $A(x)$ is compact for each $x \in X$, then there exists a Borel measurable policy (see Corollary 1 in Brown and Purves [9]).

We now make our basic assumption on the discount function δ. For a detailed discussion of various discount functions and their interpretation the reader is referred to Jaśkiewicz et al. [22].

(A1) There exists a continuous increasing function $\gamma\colon [0, \infty) \to [0, \infty)$ such that $\gamma(z) < z$ for each $z > 0$ and

$$|\delta(z_1) - \delta(z_2)| \le \gamma(|z_1 - z_2|)$$

for all $z_1, z_2 \in R$.

(A2) δ is continuous, increasing, $\delta(0) = 0$ and $\delta(-\infty) = -\infty$.

Obviously, assumption (A1) implies that $\gamma(0) = 0$.

Note that any element of D_{n+1} is of the form $(h_n, c_n) = (x_0, c_0, \ldots, x_n, c_n)$. Every function $w \in \hat{B}(D_{m+1})$ can be regarded as a function from the space $\hat{B}(H)$.

Put $U_0(x, c) = u(x, c)$ for $(x, c) \in D$. For any $(h_n, c_n) = (x_0, c_0, x_1, c_1, \ldots, x_n, c_n) \in D_{n+1}$, $n \in N$, define

$$\begin{aligned} U_n(h_n, c_n) &= u(x_0, c_0) + \delta(u(x_1, c_1) + \delta(u(x_2, c_2) + \cdots + \delta(u(x_n, c_n))\cdots)) \\ &= u(x_0, c_0) + \delta(U_{n-1}(x_1, c_1, \ldots, x_n, c_n)). \end{aligned} \tag{5}$$

Let $l > 0$ be a constant such that $u(x, c) \le l$ for all $(x, c) \in D$. Define

$$U_n^l(h) := \lim_{t\to\infty} [u(x_0, c_0) + \delta(u(x_1, c_1) + \cdots + \delta(u(x_n, c_n) + \underbrace{\delta(l + \cdots + \delta(l))}_{t \text{ terms}}))]. \tag{6}$$

From the arguments used in proving Proposition 1 in §6, it follows that the limit in (6) exists. Moreover, we can conclude the following result.

Lemma 2. *Under assumptions* (A1) *and* (A2), *for any* $h=(x_0, c_0, x_1, c_1, \dots)$,

$$U(h) := \lim_{n\to\infty} U_n(x_0, c_0, \dots, x_n, c_n) = u(x_0, c_0) + \delta(u(x_1, c_1) + \delta(u(x_2, c_2) + \cdots)) \tag{7}$$

exists in $\underline{R}$. *The sequence* $\{U_n^l(h)\}$ *is nonincreasing and*

$$\lim_{n\to\infty} U_n^l(h) = U(h), \quad h \in H.$$

Moreover, $U \in \hat{B}(H)$.

Using (7) and assumption (A2), we infer that

$$U(x_0, c_0, x_1, c_1, \dots) = u(x_0, c_0) + \delta(U(x_1, c_1, x_2, c_2, \dots)). \tag{8}$$

We call U given by (7), the *utility function* on the *sample paths*. By (8), the function U is a *recursive utility* in the sense of Koopmans [24] with the aggregator $G(a, r) = a + \delta(r)$. In the classical setup with linear discount function $\delta(r) = \beta r$, where $\beta \in (0,1)$, the utility U is of the form

$$U(h) = \sum_{t=0}^{\infty} \beta^t u(x_t, c_t)$$

introduced by Samuelson [33].

Lemma 3. *Assume* (A1) *and* (A2). *If* $u \in B(D)$, *then* $U \in B(H)$, *and* $\{U_n\}$ *converges to* U *in the space* $B(H)$ *endowed with the supremum norm.*

The proof of this result is similar to that of Theorem 1 in Jaśkiewicz et al. [22] and is based on an extension of the Banach contraction principle given in Matkowski [27].

For $n \in N_0$, define $\hat{H}^n$ as the space of sequences $h^n = (c_n, x_{n+1}, c_{n+1}, x_{n+2}, \dots)$, where $x_k \in X$, $c_k \in A$ for each $k \geq n$. For any $h_n \in H_n$, by $H^n(h_n)$ denote the set of all $h^n \in \hat{H}^n$ such that $(h_n, h^n) \in H$. It is known that $H^n(h_n)$ is a Borel subset of $\hat{H}^n$ and consists of all *feasible future histories* of the process from period n onwards.

According to the Ionescu-Tulcea theorem (see Proposition V.1.1 in Neveu [28]), for any policy $\pi \in \Pi$ and $m \in N_0$, there exists a unique conditional probability measure $P^\pi(\cdot \mid h_m)$ on $\hat{H}^m$ given $h_m \in H_m$ such that $P^\pi(H^m(h_m) \mid h_m) = 1$. Let $E^\pi_{h_m}$ denote the expectation operator corresponding to the measure $P^\pi(\cdot \mid h_m)$. For a given history $h_m \in H_m$ and $\pi \in \Pi$, we define the *expected utility from period m onwards* as follows:

$$V_m(h_m, \pi) := E^\pi_{h_m} U = \int_{\hat{H}^m} U(h_m, h^m) P^\pi(dh^m \mid h_m). \tag{9}$$

From the construction of $P^\pi(\cdot \mid h_m)$, it follows that the integral in (9) is actually taken over the set $H^m(h_m)$. By Lemma 2, $U \in \hat{B}(H)$, and therefore $V_m(h_m, \pi)$, is well defined. For each $m \in N_0$, define

$$P(h_m) := \{\nu\colon \nu = P^\pi(\cdot \mid h_m) \text{ for some } \pi \in \Pi\}$$

as the set of probability measures on the set $H^m(h_m)$ of all feasible future action-state sequences given a partial history $h_m \in H_m$. Furthermore, put

$$V_m(h_m) := \sup_{\nu \in P(h_m)} \int_{\hat{H}^m} U(h_m, h^m)\nu(dh^m) = \sup_{\pi\in\Pi} E^\pi_{h_m} U, \quad h_m \in H_m.$$

The following result is well known; see Theorem 1 in Bertsekas and Shreve [5], Theorem 3.1B in Feinberg [16] or Theorem 7.1 in Strauch [38].

Lemma 4. *For each* $m \in N_0$, *the function* V_m *is upper semianalytic.*

We now state an important consequence of Proposition 1 from §6.

Lemma 5. *For any policy* $\pi \in \Pi$, $m \in N_0$, *and* $h_m \in H_m$, *we have*

$$V_m(h_m, \pi) = E^\pi_{h_m} U = \lim_{n\to\infty} E^\pi_{h_m} U_n = \lim_{n\to\infty} E^\pi_{h_m} U_n^l.$$

Proof. Since $U_{n+1}^l \le U_n^l$ for each $n \in N_0$, then by the monotone convergence theorem and Lemma 2, we obtain that

$$\lim_{n\to\infty} E_{h_m}^{\pi} U_n^l = E_{h_m}^{\pi} U. \tag{10}$$

Moreover, from the proof of Proposition 1 in §6, it follows that for each $\epsilon > 0$ there exists $N_1 > 0$ such that for all $n > N_1$ and $h \in H$,

$$U_n^l(h) \le U_n(h) + \epsilon \le U_n^l(h) + \epsilon.$$

Therefore, $E_{h_m}^{\pi} U_n^l \le E_{h_m}^{\pi} U_n + \epsilon \le E_{h_m}^{\pi} U_n^l + \epsilon$. Combining this fact with (10), we can easily get the conclusion. □

In the sequel we shall use some operator notation which is restricted to nonrandomized policies. Let $\pi = \{\pi_n\} \in \Pi_0$. For any function $\tilde{w}$ on A we define $(\pi_k \tilde{w})(h_k) := \tilde{w}(\pi_k(h_k))$, $h_k \in H_k$, $k \in N_0$. In a similar way, for any upper semianalytic function $\hat{w}$ on H_{k+1}, we set

$$(q\hat{w})(h_k, c_k) := \int_X \hat{w}(h_k, c_k, x_{k+1}) q(dx_{k+1} \mid x_k, c_k), \quad k \in N_0.$$

By Proposition 7.48 in Bertsekas and Shreve [4], $q\hat{w}$ is also upper semianalytic. Moreover, for an upper semianalytic function w' on D_{k+1}, we put

$$(q\pi_{k+1} w')(h_k, c_k) := \int_X w'(h_k, c_k, x_{k+1}, \pi_{k+1}(h_{k+1})) q(dx_{k+1} \mid x_k, c_k).$$

We close this section with some useful remarks on the measure $P^{\pi}(\cdot \mid h_m)$. Recall that $E_{h_m}^{\pi}$ is a conditional expectation under the policy π given the history h_m. Clearly, any function w on D_{n+1} can be regarded as a function on H that depends only on the first $2(n+1)$ coordinates. Let $n > m$. Then, for an upper semianalytic function w, we have

$$E_{h_m}^{\pi} w = \int_{\hat{H}^m} w(h_m, h^m) P^{\pi}(dh^m \mid h_m) = (\pi_m q \pi_{m+1} \cdots q \pi_n w)(h_m). \tag{11}$$

Obviously, $h_m \mapsto E_{h_m}^{\pi} w$ is upper semianalytic.

Lemma 6. *Let $\pi := \{\pi_m\} \in \Pi_0$. Then*

$$\lim_{n\to\infty} (\pi_m q \pi_{m+1} q \cdots \pi_n q V_{n+1})(h_m) = E_{h_m}^{\pi} U.$$

Proof. Fix $m \in N_0$ and $h_m \in H_m$. Let $n > m$. From our definition (6), it follows that $U(h_{n+1}, h^{n+1}) \le U_n^l(h_{n+1}, h^{n+1})$ for all $h = (h_{n+1}, h^{n+1}) \in H$. Moreover, we know that the function $U_n^l(h_{n+1}, h^{n+1})$ is independent of h^{n+1}. Therefore,

$$V_{n+1}(h_{n+1}) = \sup_{\pi \in \Pi} \int_{\hat{H}^{n+1}} U(h_{n+1}, h_{n+1}) P^{\pi}(dh^{n+1} \mid h^{n+1}) \le U_n^l(h_{n+1}, h^{n+1}).$$

Thus

$$(\pi_m q \pi_{m+1} q \cdots \pi_n q V_{n+1})(h_m) \le E_{h_m}^{\pi} U_n^l.$$

Using Lemma 5, we obtain

$$\limsup_{n\to\infty} (\pi_m q \pi_{m+1} q \cdots \pi_n q V_{n+1})(h_m) \le \lim_{n\to\infty} E_{h_m}^{\pi} U_n^l = E_{h_m}^{\pi} U. \tag{12}$$

On the other hand, we have

$$(\pi_m q \pi_{m+1} q \cdots \pi_n q V_{n+1})(h_m) \ge E_{h_m}^{\pi} U. \tag{13}$$

for all $n > m$. Clearly, (12) and (13) imply the assertion. □

4. Main results. For any policy $\pi \in \Pi$ and $m \in N_0$, let $\pi^m = \{\pi_k\}_{k\geq m}$. We call π^m a *conditional policy* that can be used to select actions from stage m onwards given a partial history $h_m \in H_m$. Note that the probability measure $P^\pi(\cdot \mid h_m) \in P(h_m)$ only depends on π^m.

Definition 1. A policy $\pi^\varepsilon \in \Pi$ is called *persistently ε-optimal* for an $\varepsilon \geq 0$ if

$$E_{h_m}^{\pi^\varepsilon} U \geq \sup_{\pi\in\Pi} E_{h_m}^{\pi} U - \varepsilon$$

for each $h_m \in H_m$, $m \in N_0$. A 0-optimal policy is called *optimal*.

Note that $\pi^\varepsilon \in \Pi$ is persistently ε-optimal if the conditional policy $\pi^{m\varepsilon} = \{\pi_k^\varepsilon\}_{k\geq m}$ is ε-optimal (in the usual sense) for any partial history $h_m \in H_m$, $m \in N_0$. Persistent optimality has been widely used in gambling theory (Dubins and Savage [12], Dubins and Sudderth [13]) and is stronger than the μ-optimality studied, for example, in Schäl [35], Feinberg [15, 16] and their references (see Remark 3). As noted by Dubins and Sudderth [13], Kertz and Nachman [23] and by Schäl [36], persistently optimal policies in the nonstationary dynamic setting are counterparts of stationary optimal ones in the classical stationary models. They can be characterized by a system of Bellman equations; see Kertz and Nachman [23], Schäl [36]. Such a characterization is also used in our proof; see (15) and (16) below.

We can now state our first main result.[3]

Theorem 1. *Let the dynamic programming model satisfy assumptions* (i)–(vi). *Furthermore, assume* (A1) *and* (A2). *Then, for any $\varepsilon > 0$, there exists a nonrandomized persistently ε-optimal policy.*

Proof. By Lemma 4, the function $h_m \mapsto V_m(h_m)$ is upper semianalytic for each $m \in N_0$. In a standard way, one can show that

$$V_m(h_m) = \sup_{c\in A(x_m)} \int_X V_{m+1}(h_{m+1}) q(dx_{m+1} \mid x_m, c)$$

for every $h_m \in H_m$, $m \in N_0$. Fix $\varepsilon > 0$ and choose a sequence of positive numbers $\{\varepsilon_k\}$ such that $\sum_{k=0}^\infty \varepsilon_k = \varepsilon$. By the Yankov-von Neumann theorem (see Bertsekas and Shreve [4] or Dynkin and Yushkevich [14]), for each $m \in N_0$ and $\varepsilon_m > 0$ there exists a universally measurable function π_m^ε: $H_m \mapsto A$ such that $\pi_m^\varepsilon(h_m) \in A(x_m)$ for each $h_m \in H_m$ and

$$V_m(h_m) \leq \int_X V_{m+1}(h_{m+1}) q(dx_{m+1} \mid x_m, \pi_m^\varepsilon(h_m)) + \varepsilon_m = (\pi_m^\varepsilon q V_{m+1})(h_m) + \varepsilon_m.$$

Thus, continuation of this procedure $(n-m)$ times yields that

$$V_m(h_m) \leq (\pi_m^\varepsilon q \pi_{m+1}^\varepsilon q \cdots \pi_n^\varepsilon q V_{n+1})(h_m) + \varepsilon_m + \cdots + \varepsilon_n.$$

Put $\pi^\varepsilon = \{\pi_n^\varepsilon\}$. Letting now $n \to \infty$ and using Lemma 6, we conclude that

$$V_m(h_m) = \sup_{\pi\in\Pi} E_{h_m}^\pi U \leq \lim_{n\to\infty} (\pi_m^\varepsilon q \pi_{m+1}^\varepsilon q \cdots \pi_n^\varepsilon q V_{n+1})(h_m) + \sum_{k=m}^\infty \varepsilon_k < E_{h_m}^{\pi^\varepsilon} U + \varepsilon$$

for any $m \in N_0$ and $h_m \in H_m$, which completes the proof. □

4.1. Dynamic programming models with compact action spaces. We shall need the following two standard sets of assumptions, which will be used alternatively.

Conditions (S): The set $A(x)$ is compact for every $x \in X$, and

(S1) for each $x \in X$ and every Borel set $\tilde{X} \subset X$, the function $q(\tilde{X} \mid x, \cdot)$ is continuous on $A(x)$,

(S2) the function $u(x, \cdot)$ is upper semicontinuous on $A(x)$ for every $x \in X$.

Conditions (W): The set $A(x)$ is compact for every $x \in X$, and

(W0) the set-valued mapping $x \mapsto A(x)$ is upper semicontinuous,

(W1) the transition law q is *weakly continuous*; that is,

$$\int_X \phi(y) q(dy \mid x, c)$$

is a continuous function of $(x, c) \in D$ for each bounded continuous function ϕ,

(W2) the function u is upper semicontinuous on D.

[3] Theorem 1 is inspired by a question raised by one of the referees.

Note that Conditions (W) also embrace the case of *a deterministic optimal growth model* with a continuous production function examined in Stokey et al. [37] or Jaśkiewicz et al. [22].

Lemma 7. (a) *Let $\phi\colon X \times A \mapsto \underline{R}$ be a bounded-from-above Borel measurable function such that $a \mapsto \phi(x, a)$ is upper semicontinuous on $A(x)$ for each $x \in X$. Under condition* (S1), *the function*

$$c \mapsto \int_X \phi(y, c) q(dy \mid x, c)$$

is upper semicontinuous on $A(x)$ for every $x \in X$.

(b) *Under condition* (W1), *the function*

$$(x, c) \mapsto \int_X \phi(y, c) q(dy \mid x, c)$$

is upper semicontinuous on D for every bounded-from-above upper semicontinuous function $\phi\colon X \times A \mapsto \underline{R}$.

Proof. One can show that ϕ is a limit of a nonincreasing sequence $\{\phi_n\}$ of bounded Borel measurable functions such that every $a \mapsto \phi_n(x, a)$ is continuous on $A(x)$; see the proof of Theorem 1.2 in Nowak [29]. Then, for every $x \in X$, $c \mapsto \int_X \phi_n(y, c) q(dy \mid x, c)$ is continuous on $A(x)$ by Proposition 18 on page 231 in Royden [32]. Clearly, $\int_X \phi(y, c) q(dy \mid x, c) = \inf_{n\in N} \int_X \phi_n(y, c) q(dy \mid x, c)$, which implies part (a). Part (b) is Proposition 7.31 in Bertsekas and Shreve [4]. □

4.2. Persistently optimal policies. Approximation by truncated models. Let us first assume (S). Consider the bounded one-period utility functions,

$$u^s(x, c) := \max\{u(x, c), -s\}, \quad \text{for } (x, c) \in D, \quad s \in N.$$

All the functions defined as in §3 with u replaced by u^s will be denoted with the superscript s. Let $h_k \in H_k$, $k \in N_0$. For any $n \geq k+1$, consider a finite horizon decision problem from period k till period n with the utility function U_n^s introduced in (5) with u replaced by u^s. A policy in this model can be restricted to a sequence $\{\pi_k, \ldots, \pi_n\}$ (with π_j defined as in §3, $k \leq j \leq n$) because U_n^s depends on $(h_n, c_n) \in D_{n+1}$ only. By the backward induction, making use of Lemmas 1(a), 7(a), Conditions (S), and (11), we conclude that there exists an optimal Borel measurable policy $\{\pi_k^o, \ldots, \pi_n^o\}$, i.e.,

$$v_{k,n}^s(h_k) := \sup_{\nu \in P(h_k)} \int_{\hat{H}^k} U_n^s(h_k, h^k) \nu(dh^k) = (\pi_k^o q \pi_{k+1}^o \cdots q \pi_n^o U_n^s)(h_k)$$

for each $h_k \in H_k$. This implies that $v_{k,n}^s$ is a Borel measurable function of h_k. From the induction construction, it follows that

$$v_{m,n}^s(h_m) = \max_{c \in A(x_m)} \int_X v_{m+1,n}^s(h_{m+1}) q(dx_{m+1} \mid x_m, c), \quad h_m \in H_m, \quad m \in N_0. \tag{14}$$

(Recall that x_k is the last component of h_k.) We claim that $\{v_{m,n}^s(h_m)\}$ converges uniformly in $h_m \in H_m$ to

$$V_m^s(h_m) = \sup_{\nu \in P(h_m)} \int_{\hat{H}^m} U^s(h_m, h^m) \nu(dh^m)$$

as $n \to \infty$. Indeed, note that

$$\begin{aligned}
&\sup_{h_m \in H_m} |v_{m,n}^s(h_m) - V_m^s(h_m)| \\
&\quad = \sup_{h_m \in H_m} \left| \sup_{\nu \in P(h_m)} \int_{\hat{H}^m} U_n^s(h_m, h^m) \nu(dh^m) - \sup_{\nu \in P(h_m)} \int_{\hat{H}^m} U^s(h_m, h^m) \nu(dh^m) \right| \\
&\quad \leq \sup_{h_m \in H_m} \sup_{\nu \in P(h_m)} \int_{\hat{H}^m} |U_n^s(h_m, h^m) - U^s(h_m, h^m)| \, \nu(dh^m) \\
&\quad \leq \sup_{h_m \in H_m} \sup_{\nu \in P(h_m)} \int_{\hat{H}^m} \sup_{h = (h_m, h^m) \in H} |U_n^s(h_m, h^m) - U^s(h_m, h^m)| \, \nu(dh^m) \\
&\quad = \sup_{h \in H} |U_n^s(h) - U^s(h)|.
\end{aligned}$$

Now making use of Lemma 3, we conclude that $\sup_{h\in H}|U_n^s(h)-U^s(h)|\to 0$ as $n\to\infty$. Hence, V_m^s is bounded and Borel measurable. Using the uniform convergence showed above and (14) one can easily see that

$$V_m^s(h_m)=\max_{c\in A(x_m)}\int_X V_{m+1}^s(h_{m+1})q(dx_{m+1}\mid x_m,c) \tag{15}$$

for any $h_m\in H_m$, $m\in N_0$. This is the Bellman equation for the truncated model. Clearly, Equation (15) can also be derived under Conditions (W) with the help of Lemma 7(b).

We can now prove our second main result.

THEOREM 2. *Assume* (A1) *and* (A2). *If, additionally, the set of Conditions either* (S) *or* (W) *is satisfied, then there exists a nonrandomized Borel measurable persistently optimal policy.*

PROOF. Note that $u^s\searrow u$ as $s\to\infty$. Since δ is increasing by (A2), it follows that $\{V_m^s\}$ is nonincreasing for every $m\in N_0$. Therefore,

$$V_m^\infty(h_m):=\lim_{s\to\infty}V_m^s(h_m)$$

exists in $\underline{R}$ for every $h_m\in H_m$. Clearly, V_m^∞ is Borel measurable and $V_m^\infty(h_m)\ge V_m(h_m)$ for all $h_m\in H_m$, $m\in N_0$. Moreover, letting $s\to\infty$ in (15) and making use of Proposition 10.1 in Schäl [34] or Lemma 1 in Jaśkiewicz et al. [22] and the monotone convergence theorem, we conclude that

$$\begin{aligned}V_m^\infty(h_m)&=\lim_{s\to\infty}\max_{c\in A(x_m)}\int_X V_{m+1}^s(h_{m+1})q(dx_{m+1}\mid x_m,c)\\&=\max_{c\in A(x_m)}\lim_{s\to\infty}\int_X V_{m+1}^s(h_{m+1})q(dx_{m+1}\mid x_m,c)\\&=\max_{c\in A(x_m)}\int_X V_{m+1}^\infty(h_{m+1})q(dx_{m+1}\mid x_m,c).\end{aligned}$$

By Lemmas 1(a) and 7(a), for any $m\in N_0$, there exists a Borel measurable function $\hat{\pi}_m$ of $h_m\in H_m$ such that the maximum is attained at the point $\hat{\pi}_m(h_m)$ in the above display. Thus, we have

$$V_m^\infty(h_m)=(\hat{\pi}_m qV_{m+1}^\infty)(h_m)=\int_X V_{m+1}^\infty(h_{m+1})q(dx_{m+1}\mid x_m,\hat{\pi}_m(h_m)), \tag{16}$$

$h_m\in H_m$, $m\in N_0$. Iterating this equality, we obtain for any $m\in N_0$ and $k>m$ that

$$V_m^\infty(h_m)=(\hat{\pi}_m q\hat{\pi}_{m+1}q\cdots\hat{\pi}_k qV_{k+1}^\infty)(h_m).$$

Let $\hat{\pi}:=\{\hat{\pi}_m\}$ with $\hat{\pi}_m$ defined in (16) for each $m\in N_0$. Since δ is increasing, it follows that

$$V_m^\infty(h_m)=(\hat{\pi}_m q\hat{\pi}_{m+1}q\cdots\hat{\pi}_k qV_{k+1}^\infty)(h_m)\le(\hat{\pi}_m q\hat{\pi}_{m+1}q\cdots\hat{\pi}_k qV_{k+1}^s)(h_m) \tag{17}$$

for any $s\in N$, $h_m\in H_m$, $m\in N_0$. Now letting $k\to\infty$ in (17) and using Lemma 6 (with V_{n+1} replaced by V_{n+1}^s), we deduce that

$$V_m^\infty(h_m)\le\lim_{k\to\infty}(\hat{\pi}_m q\hat{\pi}_{m+1}q\cdots\hat{\pi}_k qV_{k+1}^s)(h_m)=E_{h_m}^{\hat{\pi}}U^s$$

for any $s\in N$, $h_m\in H_m$, $m\in N_0$. From Proposition 2 in §6, it follows that $U^s\searrow U$, as $s\to\infty$ and therefore, the monotone convergence theorem yields

$$\begin{aligned}V_m^\infty(h_m)&\le E_{h_m}^{\hat{\pi}}U=\int_{\hat{H}_m}U(h_m,h^m)P^{\hat{\pi}}(dh^m\mid h_m)\\&\le\sup_{\nu\in P(h_m)}\int_{\hat{H}^m}U(h_m,h^m)\nu(dh^m)=V_m(h_m)\end{aligned} \tag{18}$$

for each $h_m\in H_m$, $m\in N_0$. Since $V_m^\infty\ge V_m$ on H_m, from (18), we conclude that $\hat{\pi}$ is a persistently optimal policy.

The proof under Conditions (W) makes use of Lemmas 1(b), 7(b), and proceeds analogously as in the case of Conditions (S). The value functions V_m in case of Conditions (W) are upper semicontinuous. □

REMARK 1. Our proof of Theorem 2 consists of two basic steps. First, we study decision models with bounded one-period utility and use an approximation with finite horizon ones. Then, simple truncations of u are used, and monotone approximations are applied. This method is efficient, because the action spaces are compact. Counterexamples given for negative dynamic programming in Strauch [38] and Schäl [34] suggest that approximations by the truncated models are not possible if the action spaces are noncompact. However, within such a framework one can obtain a persistently ε-optimal policy (Theorem 1). The fact that we allow for u unbounded from below makes our results applicable to many models in economic theory studied in Stokey et al. [37], Jaśkiewicz and Nowak [21] and their references. Similar models with the deterministic transition function have been studied in Jaśkiewicz et al. [22], where stationary optimal policies are shown to exist. Here the problem of optimality is more involved and we cannot expect the existence of a stationary optimal policy. Finally, we would like to emphasize that our approach is also applicable for nonstationary stochastic transition functions extensively examined in Hinderer [20], Kertz and Nachman [23], Schäl [35, 36].

REMARK 2. It is worth emphasizing that the existence of persistently optimal policies under conditions similar to (W) was already shown in Theorem 5.2 of Kertz and Nachman [23]. However, their proof proceeds along different lines. They consider a more general upper semicontinuous utility function U on the space H and study some topological properties of the correspondences $h_m \mapsto P(h_m)$. In consequence, their route to the existence of a persistently optimal policy is more complex than ours. On the other hand, the result under Conditions (S) has not been clearly presented so far, and only some remarks on this issue are given in Schäl [36].

REMARK 3. Persistent optimality is essentially stronger than μ-optimality. Let us recall that a policy $\hat{\pi} \in \Pi$ is μ-optimal, if

$$V(\mu, \hat{\pi}) = \sup_{\pi \in \Pi} V(\mu, \hat{\pi}), \qquad \text{where} \qquad V(\mu, \pi) := \int_X E^{\pi}_{x_0} U \mu(dx_0),$$

and μ is a probability distribution of the initial state. The existence of nonrandomized policies under either Conditions (S) or (W) follows from Schäl [35] and Theorem 3.2 in Feinberg [16]. However, to apply the methods in Schäl [35], we still need our convergence results for the utility function U given in §6. The theorems on the existence of μ-optimal (or (μ, ε)-optimal) nonrandomized policies obtained by Feinberg [15, 16] are based upon some theorems on "decomposability of randomized strategies" studied earlier by Gihman and Skorokhod [19]. Nonetheless, the techniques used by Feinberg [16] do not imply the existence of a persistently optimal policy under Conditions (W) or (S). Namely, to select a nonrandomized policy for any partial history of the process, we would have to solve some additional measurability issues and to overcome a time consistency problem. Indeed, policies chosen for two partial histories of different length may not coincide at some future stages of the process. Moreover, even in this case the convergence results from §6 are germane. Therefore, the concept based on a study of the Bellman equations is more natural and simpler. The measure theoretic tools applied by Schäl [35] and Feinberg [16] are not needed in our studies.

5. Examples.

In the models described below with a linear discount function (standard discounting), there exist stationary optimal policies. When the discount function δ is nonlinear, an optimal policy must be more sophisticated.

EXAMPLE 1 (A STOCHASTIC OPTIMAL GROWTH MODEL). We have in mind the classical model studied in Stokey et al. [37], but with generalized discounting. Let $X = [0, s]$ be the set of all *capital stocks* where $s > 1$. If x_t is a capital stock at the beginning of period t, then consumption c_t in this period belongs to $A(x_t) := [0, x_t]$. The utility of consumption c_t is $u(c_t)$, where u: $X \mapsto \underline{R}$ is a fixed function. The evolution of the state process is described by some function f of the investment for the next period $y_t := x_t - c_t$ and some random variable ξ_t. In the literature, f is called *production technology*; see Stokey et al. [37]. We shall view this model as a decision process with $X = [0, s]$, $A(x) = [0, x]$, and $u(x, c) = u(c)$, $x \in X$, $c \in A(x)$. Suppose that $\{\xi_t\}$ are independent and have a common probability distribution μ with support included in $[0, z]$ for some $z > 1$. Assume that

$$x_{t+1} = f(x_t - c_t)\xi_t, \quad \text{for } t = 0, 1, \ldots,$$

where f: $X \mapsto R$ is a continuous and increasing function, $f(0) = 0$,

$$(0, s) \ni y \to \frac{f(y)}{y} \quad \text{is strictly decreasing;} \tag{19}$$

$$\lim_{y \to 0+} \frac{f(y)}{y} > 1 \qquad \text{and} \qquad \frac{f(s)}{s} < 1. \tag{20}$$

Conditions (19) and (20) imply that there exists $y_0 > 0$ such that

$$f(y) > y \quad \text{for all } y \in (0, y_0), \qquad \text{and} \qquad f(y) < y \quad \text{for all } y \in (y_0, s).$$

We assume that $f(s)z \le s$. Then for any $x_t \in X$, $c_t \in A(x_t)$, and $\xi_t \in [0, z]$,

$$x_{t+1} = f(x_t - c_t)\xi_t \in X.$$

Observe that the transition probability q is of the form

$$q(B \mid x, c) = \int_0^s 1_B(f(x-c)\xi)\mu(d\xi),$$

where $B \subset X$ is a Borel set, $x \in X$, $c \in A(x)$. Here, 1_B is the indicator function of the set B. If v is a continuous function on X, then the integral

$$\int_X v(y)q(dy \mid x, c) = \int_0^s v(f(x-c)\xi)\mu(d\xi)$$

depends continuously on (x, c). This example allows for $u(c) = \log c$ as a one-period utility function where $\log 0 = -\infty$. For any nonlinear discount function δ satisfying (A1) and (A2), there exists a persistently optimal policy.

Example 2 (An Inventory Model). A manager sells a certain amount of goods each period $t = 0, 1, \ldots$ at price p. If he has $x_t \ge 0$ units in stock, he can sell $\min\{x_t, D_t\}$, where $D_t \ge 0$ is a *continuous* random variable representing an unknown demand. He can also order any amount c_t of new goods to be delivered at the beginning of the next period at a cost $z(c_t)$ paid immediately. It is assumed that z is continuous, increasing, and $z(0) = 0$. The state equation is of the form:

$$x_{t+1} = x_t - \min\{x_t, D_t\} + c_t, \quad \text{for } t = 0, 1, \ldots,$$

where $\{D, D_t\}$ is a sequence of i.i.d. random variables such that D follows a distribution F with $ED < \infty$. The manager discounts his revenues according to a function δ satisfying (A1) and (A2). This model can be viewed as a decision process, in which $X := [0, \infty)$ is the state space (i.e., the set of possible levels of stock), $A = A(x) := [0, K]$ is the action space, where $K > 0$, and $u(x, c) := Ep\min\{x, D\} - z(c)$ is the one period utility or return function.

Clearly, $u(x, c) \le pED$. Next note that the transition probability q is of the form

$$q(B \mid x, c) = \int_0^\infty 1_B(x - \min\{x, y\} + c)\, dF(y),$$

where $B \subset X$ is a Borel set, $x \in X$, $c \in A$. If v is a bounded continuous function on X, then the integral

$$\begin{aligned}\int_X v(y)q(dy \mid x, c) &= \int_0^\infty v(x - \min\{x, y\} + c)\, dF(y)\\ &= \int_0^x v(x - y + c)\, dF(y) + \int_x^\infty v(c)\, dF(y)\\ &= \int_0^x v(x - y + c)\, dF(dy) + v(c)(1 - F(x))\end{aligned}$$

depends continuously on (x, c). Hence, the model satisfies Conditions (W). Therefore, for any nonlinear discount function δ satisfying (A1) and (A2), there exists a persistently optimal policy.

6. Basic convergence results. Let assumptions (A1) and (A2) be satisfied. By $\{r_n\}_0^\infty$ we denote a sequence such that $r_n \in \underline{R}$ and $r_n \le l$ for each $n \ge 0$ and some $l > 0$. Define the following functions:

$$w_0(r_0) = r_0, \qquad w_n(r_0, r_1, \ldots, r_n) = r_0 + \delta(w_{n-1}(r_1, \ldots, r_n)) \quad \text{for } n \ge 1.$$

Note that w_n defined above is a function of $n+1$ variables. We first use the functions w_n for the sequence $\{r_n\}_0^\infty$ with $r_n = l$ for all $n \ge 0$. Put $l_{n+1} := (r_0, \ldots, r_n)$ if $r_t = l$ for $t = 0, \ldots, n$, $n \ge 0$.

In the proof of our first result we use a simple argument from Cho and O'Regan [10].

Lemma 8. *There exists* $L := \lim_{m\to\infty} w_m(l_{m+1}) = \sup_{m\ge 1} w_m(l_{m+1}) < \infty$.

PROOF. Note that (since the function δ is increasing), for each $m \geq 1$,

$$w_m(l_{m+1}) \geq w_{m-1}(l_m).$$

Hence, the sequence $\{w_m(l_{m+1})\}$ is nondecreasing. We show that its limit is finite. Indeed, observe that by (A1),

$$w_1(l,l) - w_0(l) = l + \delta(l) - l \leq \gamma(l),$$

$$w_2(l,l,l) - w_1(l,l) = l + \delta(w_1(l,l)) - l - \delta(w_0(l)) \leq \gamma(w_1(l,l) - w_0(l)) \leq \gamma^{(2)}(l),$$

where $\gamma^{(m)}$ denotes the composition of γ with itself m times ($m \geq 2$). Continuing this procedure one can see that

$$w_m(l_{m+1}) - w_{m-1}(l_m) \leq \gamma^{(m)}(l).$$

Let $\epsilon > 0$ be fixed. Since $\gamma^{(m)}(l) \to 0$ as $m \to \infty$, there exists $m \geq 1$ such that

$$w_m(l_{m+1}) - w_{m-1}(l_m) \leq \epsilon - \gamma(\epsilon).$$

Note now that

$$\begin{aligned} w_{m+1}(l_{m+2}) - w_{m-1}(l_m) &= w_{m+1}(l_{m+2}) - w_m(l_{m+1}) + w_m(l_{m+1}) - w_{m-1}(l_m) \\ &\leq l + \delta(w_m(l_{m+1})) - l - \delta(w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon) \\ &\leq \gamma(w_m(l_{m+1}) - w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon) \\ &\leq \gamma(\epsilon - \gamma(\epsilon)) + \epsilon - \gamma(\epsilon) \leq \gamma(\epsilon) + \epsilon - \gamma(\epsilon) = \epsilon. \end{aligned}$$

Similarly,

$$\begin{aligned} w_{m+2}(l_{m+3}) - w_{m-1}(l_m) &= w_{m+2}(l_{m+3}) - w_m(l_{m+1}) + w_m(l_{m+1}) - w_{m-1}(l_m) \\ &\leq l + \delta(w_{m+1}(l_{m+2})) - l - \delta(w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon) \\ &\leq \gamma(w_{m+1}(l_{m+2}) - w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon) \\ &\leq \gamma(\epsilon) + \epsilon - \gamma(\epsilon) = \epsilon. \end{aligned}$$

Thus, by induction we obtain that

$$w_{m+k}(l_{m+k+1}) - w_{m-1}(l_m) \leq \epsilon$$

for all $k \geq 0$. Hence, $w_{m+k}(l_{m+k+1}) \leq w_{m-1}(l_m) + \epsilon$. Since $w_{m-1}(l_m)$ is finite, it follows that L is finite. □

Consider a sequence $\{u_n\}_0^\infty$ of utilities $u_n \in \underline{R}$ such that $u_n \leq l$ for each $n \geq 0$.

For $m \geq 1$ and $n \geq 0$, let us introduce the following notation:

$$W_{n,m}(u_0, \ldots, u_n, l_m) := w_{n+m}(u_0, \ldots, u_n, l_m).$$

Note that $W_{n,m}$ is a function on a $(n+m+1)$-dimensional vector. For example,

$$W_{2,3}(u_0,u_1,u_2,l_3) = W_{2,3}(u_0,u_1,u_2,l,l,l) = u_0 + \delta(u_1 + \delta(u_2 + \delta(l + \delta(l + \delta(l))))),$$

$$W_{3,2}(u_0,u_1,u_2,u_3,l_2) = W_{4,2}(u_0,u_1,u_2,u_3,l,l) = u_0 + \delta(u_1 + \delta(u_2 + \delta(u_3 + \delta(l + \delta(l))))).$$

PROPOSITION 1. *The limit of the sequence $\{w_n(u_0, \ldots, u_n)\}$ exists in $\underline{R}$.*

PROOF. We first study the case where $u_n > -\infty$ for all $n \geq 0$. Note that for each $n \geq 0$ and $m \geq 1$

$$w_n(u_0, \ldots, u_n) \leq W_{n,m}(u_0, \ldots, u_n, l_m). \tag{21}$$

Moreover,

$$\begin{aligned} &W_{n,m}(u_0, \ldots, u_n, l_m) - w_n(u_0, \ldots, u_n) \\ &\quad = u_0 + \delta(W_{n-1,m}(u_1, \ldots, u_n, l_m)) - u_0 - \delta(w_{n-1}(u_1, \ldots, u_n)) \\ &\quad \leq \gamma(W_{n-1,m}(u_1, \ldots, u_n, l_m) - w_{n-1}(u_1, \ldots, u_n)). \end{aligned}$$

Similarly,

$$\begin{aligned}&W_{n-1,m}(u_1,\dots,u_n,l_m)-w_{n-1}(u_1,\dots,u_n)\\&\quad=u_0+\delta(W_{n-2,m}(u_2,\dots,u_n,l_m))-u_0-\delta(w_{n-2}(u_2,\dots,u_n))\\&\quad\le\gamma(W_{n-2,m}(u_2,\dots,u_n,l_m)-w_{n-2}(u_2,\dots,u_n)).\end{aligned}$$

Thus,

$$W_{n,m}(u_0,\dots,u_n,l_m)-w_n(u_0,\dots,u_n)\le\gamma^{(2)}(W_{n-2,m}(u_2,\dots,u_n,l_m)-w_{n-2}(u_2,\dots,u_n)).$$

Continuing in this way and using Lemma 8, we obtain

$$\begin{aligned}W_{n,m}(u_0,\dots,u_n,l_m)-w_n(u_0,\dots,u_n)&\le\gamma^{(n)}(W_{0,m}(u_n,l_m)-w_0(u_n))\\&=\gamma^{(n)}(\delta(w_{m-1}(l_m)))\le\gamma^{(n+1)}(L),\end{aligned}$$

for all $m\ge1$. Let $\epsilon>0$ be fixed. Then, for sufficiently large n, say $n>N_1$,

$$W_{n,m}(u_0,\dots,u_n,l_m)\le w_n(u_0,\dots,u_n)+\epsilon \tag{22}$$

for all $m\ge1$. Clearly, for any $m\ge1$, we have

$$\begin{aligned}W_{n,m}(u_0,\dots,u_n,l_m)&\le W_{n,m+1}(u_0,\dots,u_n,l_{m+1})\\&\le w_{n+m+1}(l_{n+m+2})\le\sup_{t\ge1}w_t(l_{t+1})=L<\infty.\end{aligned}$$

Therefore, $\lim_{m\to\infty}W_{n,m}(u_0,\dots,u_n,l_m)$ exists and is bounded from above by L. Let us denote this limit by G_n. From (21) and (22) we conclude that

$$G_n-\epsilon\le w_n(u_0,\dots,u_n)\le G_n$$

for $n>N_1$. Observe that $\{G_n\}$ is decreasing and $G_*:=\lim_{n\to\infty}G_n$ exists in $\underline{R}$. Hence, the limit

$$\lim_{n\to\infty}w_n(u_0,\dots,u_n)$$

also exists and equals G_*. Assume now that $u_n=-\infty$ for some $n\ge0$. Then

$$w_n(u_0,\dots,u_n)=-\infty$$

and

$$W_{n,m}(u_0,\dots,u_n,l_m)=-\infty$$

for all $m\ge1$. Therefore,

$$G_n=-\infty\qquad\text{and}\qquad\lim_{n\to\infty}w_n(u_0,\dots,u_n)=G_*=\lim_{n\to\infty}G_n=-\infty.\quad\square$$

For $k\ge1$, let us define $u_n^k:=\max\{u_n,-k\}$. We have arrived at our final result in this section.

Proposition 2. $\lim_{n\to\infty}w_n(u_0,\dots,u_n)=\inf_{k\ge1}\lim_{n\to\infty}w_n(u_0^k,\dots,u_n^k)$.

Proof. Assume first that $\lim_{n\to\infty}w_n(u_0,\dots,u_n)>-\infty$. Let $\epsilon>0$ be fixed. Then, by (22), we have

$$w_n(u_0,\dots,u_n)\ge W_{n,m}(u_0,\dots,u_n,l_m)-\epsilon$$

for all $m\ge1$ and $n>N_1$. Moreover, there exists N_2 such that for all $n>N_2$,

$$\lim_{t\to\infty}w_t(u_0,\dots,u_t)\ge w_n(u_0,\dots,u_n)-\epsilon. \tag{23}$$

Let us now fix $n>\max\{N_1,N_2\}$. Since $u_i^k\to u_i$ for $i\in\{0,\dots,n\}$ as $k\to\infty$, then there exists $K_1>0$ such that for $k>K_1$,

$$u_i^k-u_i\le\frac{\epsilon}{n+1}\quad\text{for each } i\in\{0,\dots,n\}.$$

Fix any $k > K_1$. By assumption (A1) we obtain that

$$\begin{aligned} w_n(u_0^k, \ldots, u_n^k) - w_n(u_0, \ldots, u_n) &\leq u_0^k - u_0 + \gamma(w_{n-1}(u_1^k, \ldots, u_n^k) - w_{n-1}(u_1, \ldots, u_n)) \\ &\leq \frac{\epsilon}{n+1} + w_{n-1}(u_1^k, \ldots, u_n^k) - w_{n-1}(u_1, \ldots, u_n). \end{aligned}$$

Similarly,

$$\begin{aligned} w_{n-1}(u_1^k, \ldots, u_n^k) - w_{n-1}(u_1, \ldots, u_n) &\leq u_1^k - u_1 + \gamma(w_{n-2}(u_2^k, \ldots, u_n^k) - w_{n-2}(u_2, \ldots, u_n)) \\ &\leq \frac{\epsilon}{n+1} + w_{n-2}(u_2^k, \ldots, u_n^k) - w_{n-2}(u_2, \ldots, u_n). \end{aligned}$$

Hence

$$w_n(u_0^k, \ldots, u_n^k) - w_n(u_0, \ldots, u_n) \leq \frac{2\epsilon}{n+1} + w_{n-2}(u_2^k, \ldots, u_n^k) - w_{n-2}(u_2, \ldots, u_n).$$

Proceeding along this line, we finally obtain

$$w_n(u_0^k, \ldots, u_n^k) - w_n(u_0, \ldots, u_n) \leq \frac{n\epsilon}{n+1} + w_0(u_n^k) - w_0(u_n) = \frac{n\epsilon}{n+1} + u_n^k - u_n \leq \epsilon.$$

Similarly, for any $m \geq 1$, we have that

$$\begin{aligned} &W_{n,m}(u_0^k, \ldots, u_n^k, l_m) - W_{n,m}(u_0, \ldots, u_n, l_m) \\ &\quad = w_{n+m}(u_0^k, \ldots, u_n^k, l_m) - w_{n+m}(u_0, \ldots, u_n, l_m) \leq \epsilon. \end{aligned}$$

Hence, we infer that

$$\begin{aligned} w_n(u_0, \ldots, u_n) &\geq W_{n,m}(u_0, \ldots, u_n, l_m) - \epsilon \\ &\geq W_{n,m}(u_0^k, \ldots, u_n^k, l_m) - 2\epsilon \\ &= w_{n+m}(u_0^k, \ldots, u_n^k, l_m) - 2\epsilon \\ &\geq w_{n+m}(u_0^k, \ldots, u_n^k, u_{n+1}^k, \ldots, u_{n+m}^k) - 2\epsilon. \end{aligned} \tag{24}$$

By Proposition 1, we deduce that

$$\lim_{m\to\infty} w_{n+m}(u_0^{k'}, \ldots, u_n^{k'}, u_{n+1}^{k'}, \ldots, u_{n+m}^{k'}) =: G_*^{k'}$$

exists for any positive integer k'. Therefore, by (24), we have

$$w_n(u_0, \ldots, u_n) \geq G_*^k - 2\epsilon, \quad k > K_1.$$

Now applying (23) we obtain that

$$\lim_{t\to\infty} w_t(u_0, \ldots, u_t) \geq G_*^k - 3\epsilon \geq \inf_{k'\geq 1} G_*^{k'} - 3\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we get that

$$\lim_{t\to\infty} w_t(u_0, \ldots, u_t) \geq \inf_{k'\geq 1} G_*^{k'}.$$

On the other hand, it is obvious that

$$\lim_{t\to\infty} w_t(u_0, \ldots, u_t) \leq \inf_{k\geq 1} \lim_{t\to\infty} w_t(u_0^k, \ldots, u_t^k) = \inf_{k\geq 1} G_*^k.$$

Combining the last two inequalities we get the conclusion.

If $\lim_{n\to\infty} w_n(u_0, \ldots, u_n) = -\infty$, then for any $M < 0$ there exists N_3 such that for all $n > N_3$,

$$M > w_n(u_0, \ldots, u_n).$$

Proceeding as above we obtain that $\inf_{k\geq 1} G_*^k = -\infty$. If $u_n = -\infty$ for some $n \geq 0$, then the proof that

$$\inf_{k\geq 1} \lim_{n\to\infty} w_n(u_0^k, \ldots, u_n^k) = \lim_{n\to\infty} w_n(u_0, \ldots, u_n) = -\infty,$$

is simple. □

Acknowledgments. The authors wish to thank the referees for their constructive comments concerning both style and content. This work is partially supported by the National Science Centre [Grant DEC-2011/03/B/ST1/00325].

References

[1] Becker RA, Boyd III, JH (1997) *Capital Theory, Equilibrium Analysis and Recursive Utility* (Blackwell Publishers, New York).
[2] Berge C (1963) *Topological Spaces* (Macmillan, New York).
[3] Bertsekas DP (1977) Monotone mappings with application in dynamic programming. *SIAM J. Control Optim.* 15(3):438–464.
[4] Bertsekas DP, Shreve SE (1978) *Stochastic Optimal Control: The Discrete-Time Case* (Academic Press, New York).
[5] Bertsekas DP, Shreve SE (1979) Universally measurable policies in dynamic programming. *Math. Oper. Res.* 4(1):15–30.
[6] Blackwell D (1965) Discounted dynamic programming. *Ann. Math. Statist.* 36(1):226–235.
[7] Boyd III, JH (1990) Recursive utility and the Ramsey problem. *J. Econ. Theory* 50(2):326–345.
[8] Brock WA, Mirman LJ (1972) Optimal economic growth and uncertainty: The discounted case. *J. Econ. Theory* 4(3):479–513.
[9] Brown LD, Purves R (1973) Measurable selections of extrema. *Ann. Statist.* 1(5):902–912.
[10] Cho Y, O'Regan D (2008) Fixed point theory for Volterra contractive operators of Matkowski type in Fréchet spaces. *Dyn. Contin. Discrete Impuls. Syst. Ser. A Math. Anal.* 15(6):871–884.
[11] Denardo EV (1967) Contraction mappings in the theory underlying dynamic programming. *SIAM Rev.* 9(2):165–177.
[12] Dubins LE, Savage LJ (1976) *Inequalities for Stochastic Processes* (*How to Gamble if You Must*) (Dover, New York).
[13] Dubins LE, Sudderth WD (1977) Persistently ϵ-optimal strategies. *Math. Oper. Res.* 2(2):125–134.
[14] Dynkin EB, Yushkevich AA (1979) *Controlled Markov Processes* (Springer, Berlin, New York).
[15] Feinberg EA (1982) Nonrandomized Markov and semi-Markov strategies in dynamic programming. *Theory Probab. Appl.* 27(1): 116–126.
[16] Feinberg EA (1982) Controlled Markov processes with arbitrary numerical criteria. *Theory Probab. Appl.* 27(3):486–503.
[17] Feinberg EA (2002) Total reward criteria. Feinberg EA, Shwartz A, eds. *Handbook of Markov Decision Processes: Theory and Methods* (Kluwer Academic Publishers, Dordrecht, The Netherlands), 173–208.
[18] Feinberg EA, Shwartz A, eds. (2002) *Handbook of Markov Decision Processes: Theory and Methods* (Kluwer Academic Publishers, Dordrecht, The Netherlands).
[19] Gihman II, Skorokhod AV (1979) *Controlled Stochastic Processes* (Springer, New York).
[20] Hinderer K (1970) *Foundations of Non-Stationary Dynamic Programming with Discrete-Time Parameter*, Lecture Notes in Operations Research, Vol. 33 (Springer-Verlag, New York).
[21] Jaśkiewicz A, Nowak AS (2011) Stochastic games with unbounded payoffs: Applications to robust control in economics. *Dyn. Games Appl.* 1(2):253–279.
[22] Jaśkiewicz A, Matkowski J, Nowak AS (2011) On variable discounting in dynamic programming: Applications to resource extraction and other economic models. *Ann. Oper. Res.* in press.
[23] Kertz RP, Nachman DC (1979) Persistently optimal plans for nonstationary dynamic programming: The topology of weak convergence case. *Ann. Probab.* 7(5):811–826.
[24] Koopmans TC (1960) Stationary ordinal utility and impatience. *Econometrica* 28(2):287–309.
[25] Kreps DM, Porteus EL (1978) Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46(1):185–200.
[26] Lucas Jr, RE, Stokey N (1984) Optimal growth with many consumers. *J. Econ. Theory* 32(1):139–171.
[27] Matkowski J (1975) Integral solutions of functional equations. *Dissertationes Math.* 127:1–68.
[28] Neveu J (1965) *Mathematical Foundations of the Calculus of Probability* (Holden-Day, San Francisco).
[29] Nowak AS (1985) Measurable selection theorems for minimax stochastic optimization problems. *SIAM J. Control Optim.* 23(3): 466–476.
[30] Nowak AS (1986) Semicontinuous nonstationary stochastic games. *J. Math. Anal. Appl.* 117(1):84–99.
[31] Porteus E (1982) Conditions for characterizing the structure of optimal strategies in infinite-horizon dynamic programs. *J. Optim. Theory Appl.* 36(3):419–431.
[32] Royden H (1968) *Real Analysis* (Macmillan, New York).
[33] Samuelson P (1937) A note on measurement of utility. *Rev. Econ. Stud.* 4(2):155–161.
[34] Schäl M (1975) Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. *Z. Wahrsch. Verw. Gebiete* 32(3):179–196.
[35] Schäl M (1975) On dynamic programming: Compactness of the space of policies. *Stochastic Process. Appl.* 3(4):345–364.
[36] Schäl M (1981) Utility functions and optimal policies in sequential decision problems. Moeschlin O, Pallaschke D, eds. *Game Theory and Mathematical Economics* (North-Holland, Amsterdam), 357–365.
[37] Stokey NL, Lucas Jr, RE, Prescott E (1989) *Recursive Methods in Economic Dynamics* (Harvard University Press, Cambridge, MA).
[38] Strauch R (1966) Negative dynamic programming. *Ann. Math. Statist.* 37(4):871–890.
[39] Weil P (1990) Nonexpected utility in macroeconomics. *Quart. J. Econom.* 105(1):29–42.