Stochastic Dynamic Programming with Non-linear Discounting

Nicole Bäuerle[1] · Anna Jaśkiewicz[2] · Andrzej S. Nowak[3]

Abstract

In this paper, we study a Markov decision process with a non-linear discount function and with a Borel state space. We define a recursive discounted utility, which resembles non-additive utility functions considered in a number of models in economics. Non-additivity here follows from non-linearity of the discount function. Our study is complementary to the work of Jaśkiewicz et al. (Math Oper Res 38:108–121, 2013), where also non-linear discounting is used in the stochastic setting, but the expectation of utilities aggregated on the space of all histories of the process is applied leading to a non-stationary dynamic programming model. Our aim is to prove that in the recursive discounted utility case the Bellman equation has a solution and there exists an optimal stationary policy for the problem in the infinite time horizon. Our approach includes two cases: (*a*) when the one-stage utility is bounded on both sides by a weight function multiplied by some positive and negative constants, and (*b*) when the one-stage utility is unbounded from below.

Keywords Stochastic dynamic programming · Non-linear discounting · Bellman equation · Optimal stationary policy

Mathematics Subject Classification 90C40 · 60J05 · 90C39 · 90B32 · 90B62

✉ Anna Jaśkiewicz
anna.jaskiewicz@pwr.edu.pl

Nicole Bäuerle
nicole.baeuerle@kit.edu

Andrzej S. Nowak
a.nowak@wmie.uz.zgora.pl

[1] Department of Mathematics, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2] Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wyspiańskiego 27, 50-370 Wrocław, Poland

[3] Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Licealna 9, 65-417 Zielona Gora, Poland

1 Introduction

Discounting future benefits and costs is crucial in order to determine fair prices for investments and projects, in particular, when long time horizons come into play like for example in the task to price carbon emissions. The idea behind pricing carbon emissions is to add up the cost of the economic damage caused by this emission for the society. Up to now the value of one ton CO_2 emissions varies quite heavily between countries, ranging for instance from \$119.43 in Sweden to \$2.69 in Japan (see [11]). This is of course only partly due to the use of different discount functions, but nevertheless emphasises the role of discounting.

The traditional discounting with a constant discount factor can be traced back to [29]. This is still the most common way to discount benefits and costs, in particular because it simplifies the computation. However, Koopmans [25] gave an axiomatic characterisation of a class of recursive utilities, which also includes the classical way of discounting. He introduced an aggregator W to aggregate the current utility u_t with future ones v_{t+1} in $v_t = W(u_t, v_{t+1})$. When we choose $W(u, v) = u + \beta v$, we get the classical discounting with a discount factor β.

In this paper, we study a Markov decision process with a Borel state space, unbounded stage utility and with a non-linear discount function δ which has certain properties. We use an aggregation of the form

$$v_t(h_t) = u_t(\pi_t(h_t)) + \int \delta\big(v_{t+1}(h_t, \pi_t(h_t), x_{t+1})\big) q(dx_{t+1}|x_t, \pi_t(h_t))$$

where q is the transition kernel of the Markov decision process, π_t is the decision function at time t and h_t the history of the process. When $\delta(x) = \beta x$ we are back in the classical setting. In this case, it is well-known how to solve a Markov decision process with an infinite time horizon, see for example [3,8,15,16,30,31]. In the unbounded utility case, the established method is to use a weighted supremum norm and combine it with Banach's contraction theorem (see e.g. [3,4,9,14,17]). In our setting the Banach contraction principle cannot be applied. Indeed our paper is in the spirit of [20–22], where also non-linear discounting was used and an extension of the Banach theorem due to Matkowski [26] was applied. Whereas papers [21,22] consider a purely deterministic decision process, work [20] treats a stochastic problem. However, in [20] the expectation is the final operator applied at the end of aggregation, whereas in the present paper expectation and discounting operators alternate. As will be explained in Sect. 3 this has the advantage that we get optimal stationary policies in our setting.

The main result of our paper is a solution procedure of these new kind of discounting problems with stochastic transition. In particular, we provide an optimality equation, show that it has a solution and prove that there exists an optimal stationary policy for the problem in the infinite time horizon. Note that we allow the utility function to be unbounded.

The outline of our paper is as follows. We introduce our model data together with the assumptions in Sect. 2. In Sect. 3, we present our optimisation problem. Particularly, we explain how the utility is aggregated in our model and what precisely the difference to the model and results in [20] is. In Sect. 4, we summarise some auxiliary results like

a measurable selection theorem and a generalised fixed point theorem, which is used later. Next, in Sect. 5.1 we treat the model, where the positive and negative part of the one-stage utility is bounded by a weight function ω. We show in this case that the value function v^* is a unique fixed point of the corresponding maximal reward operator and that every maximiser in the Bellman equation for v^* defines an optimal stationary policy. In Sect. 5.2, we consider then the setting, where the utility function is unbounded from below, but still bounded from above by the weight function ω. Here, we can only show that the value function $\underline{v}^*$ is a fixed point of the corresponding maximal reward operator, but examples show that the fixed point is not necessarily unique. Anyway, as in Sect. 5.1, any maximiser in the Bellman equation for $\underline{v}^*$ defines again an optimal stationary policy. The proof employs an approximation of $\underline{v}^*$ by a monotone sequence of value functions, which are bounded by a weight function ω in absolute value like in Sect. 5.1. In Sect. 6, we briefly discuss two numerical algorithms for the solution of problems in Sect. 5.1, namely policy iteration and policy improvement. The last section presents some applications. We discuss two different optimal growth models, an inventory problem and a stopping problem.

2 The Dynamic Programming Model

Let $\mathbb{N}$ ($\mathbb{R}$) denote the set of all positive integers (all real numbers) and $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$, $\mathbb{R}_+ = [0, \infty)$. A *Borel space* Y is a non-empty Borel subset of a complete separable metric space. By $\mathcal{B}(Y)$ we denote the σ-algebra of all Borel subsets of Y and we write $\mathcal{M}(Y)$ to denote the set of all Borel measurable functions $g : Y \to \overline{\mathbb{R}}$.

A *discrete-time Markov 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 non-empty Borel subset of $X \times A$. We assume that for each $x \in X$, the non-empty x-section

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

  of D represents the set of *actions available in state x*.
- (iv) q is a *transition probability* from D to X. For each $B \in \mathcal{B}(X)$, $q(B|x, a)$ is the probability that the new state is in the set B, given the current state is $x \in X$ and an action $a \in A(x)$ has been chosen.
- (v) $u \in \mathcal{M}(D)$ is a *one-period utility function*.
- (vi) $\delta : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ is a *discount function*.

Let $D^n := D \times \cdots \times D$ (n times) for $n \in \mathbb{N}$. Let $H_1 = X$ and H_{n+1} be the space of *admissible histories* up to the n-th transition, i.e., $H_{n+1} := D^n \times X$ for $n \in \mathbb{N}$. An element of H_n is called a partial history of the process. We put $H = D \times D \times \cdots$ and assume that H_n and H are equipped with the product σ-algebras.

In this paper, we restrict ourselves to deterministic policies, since randomisation does not give any advantage from the point of view of utility maximisation. A *policy π* is a sequence (π_n) of decision rules where, for every $n \in \mathbb{N}$, π_n is a Borel measurable

mapping, which associates any admissible history $h_n \in H_n$ ($n \in \mathbb{N}$) with an action $a_n \in A(x_n)$. We write Π to denote the *set of all policies*. Let F be the set of all Borel measurable mappings $f : X \to A$ such that $f(x) \in A(x)$ for every $x \in X$. When $A(x)$ is compact for each $x \in X$, then from the Arsenin-Kunugui result (see Theorem 18.18 in [24]), it follows that $F \neq \emptyset$. A policy π is called *stationary* if $\pi_n = f$ for all $n \in \mathbb{N}$ and some $f \in F$. Therefore, a stationary policy $\pi = (f, f, \ldots)$ will be identified with $f \in F$ and the *set of all stationary policies* will be denoted by F.

2.1 Assumptions with Comments

Let $\omega : X \to [1, \infty)$ be a fixed Borel measurable function.
Assumptions (A)

$(A2.1)$ there exists $b > 0$ such that

$$u(x, a) \geq -b\omega(x) \quad \text{for all } (x, a) \in D,$$

$(A2.2)$ there exists $c > 0$ such that

$$u(x, a) \leq c\omega(x) \quad \text{for all } (x, a) \in D.$$

Our next assumptions are on the discount function δ.

Assumptions (B)

$(B2.1)$ there exists an increasing function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\gamma(z) < z$ for each $z > 0$ and

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

for all $z_1, z_2 \in \mathbb{R}$.
$(B2.2)$ δ is increasing, $\delta(0) = 0$ and $\delta(-\infty) = -\infty$.
$(B2.3)$ (i) γ is subadditive, i.e., $\gamma(y + z) \leq \gamma(y) + \gamma(z)$ for all $y, z \geq 0$, and
(ii) $\gamma(\omega(x)y) \leq \omega(x)\gamma(y)$ for all $x \in X$ and $y > 0$.
$(B2.4)$ it holds that

$$\int_X \omega(y)q(dy|x, a) \leq \alpha\omega(x) \text{ for all } (x, a) \in D$$

and either

(i) $\alpha \leq 1$ or
(ii) $\alpha > 1$ and $\alpha\gamma(x) < x$ for all $x \in (0, +\infty)$.

Remark 2.1 In some empirical studies it was observed that negative and positive utilities were discounted by different discount factors ("sign effect"). Therefore a simple non-linear discount function is $\delta(y) = \delta_1 y$ for $y \leq 0$ and $\delta(y) = \delta_2 y$ for $y > 0$, where $\delta_1, \delta_2 \in (0, 1)$ and $\delta_1 \neq \delta_2$. For a discussion and interpretation of this and other types of discount functions the reader is referred to Jaśkiewicz, Matkowski and Nowak [21]. Additional examples of discount functions are also given in Sect. 7.

Remark 2.2 Obviously, assumption $(B2.1)$ implies that $\gamma(0) = 0$ and γ is continuous at zero. Moreover, it implies that δ is continuous and (with $(B2.2)$) that $|\delta(z)| \leq \gamma(|z|)$ for all $z \in \mathbb{R}$. From $(B2.3)$, it follows that

$$|\gamma(y) - \gamma(z)| \leq \gamma(|y - z|) \quad \text{for all} \quad y, \ z \geq 0. \tag{2.1}$$

This fact and continuity of γ at zero, implies that γ is continuous at any point in $\mathbb{R}_+$.

Remark 2.3 Assumption $(B2.3)$ holds if the function $z \mapsto \gamma(z)/z$ is non-increasing on $(0, \infty)$. Note that under this condition we have $\frac{\gamma(y+z)}{y+z} \leq \frac{\gamma(y)}{y}$ and $\frac{\gamma(y+z)}{y+z} \leq \frac{\gamma(z)}{z}$. Hence, $\gamma(y + z) \leq \gamma(y) + \gamma(z)$. Moreover, for any $d \geq 1$ and $z > 0$, $\gamma(dz)/dz \leq \gamma(z)/z$. Thus $\gamma(dz) \leq d\gamma(z)$. Take $d = \omega(x)$. Then (ii) in $(B2.3)$ holds.

Remark 2.4 There are subadditive functions γ such that $z \mapsto \gamma(z)/z$ is not necessarily non-increasing. An example of such a subadditive function is $\gamma(x) = (1 - \varepsilon)x + \varepsilon|\sin x|$ for some $\varepsilon \in (0, 1)$.

The following two standard sets of assumptions will be used alternatively.

Assumptions (W)

$(W2.1)$ $A(x)$ is compact for every $x \in X$ and the set-valued mapping $x \mapsto A(x)$ is upper semicontinuous, i.e., $\{x \in X : \ A(x) \cap K \neq \emptyset\}$ is closed for each closed set $K \subset A$,

$(W2.2)$ the function u is upper semicontinuous on D,

$(W2.3)$ the transition probability q is weakly continuous, i.e.,

$$(x, a) \mapsto \int_X \phi(y) q(dy|x, a)$$

is continuous on D for each bounded continuous function ϕ,

$(W2.4)$ the function ω is continuous on X.

$(W2.5)$ the function

$$(x, a) \mapsto \int_X \omega(y) q(dy|x, a)$$

is continuous on D.

Assumptions (S)

$(S2.1)$ $A(x)$ is compact for every $x \in X$,

$(S2.2)$ the function $u(x, \cdot)$ is upper semicontinuous on $A(x)$ for every $x \in X$,

$(S2.3)$ for each $x \in X$ and every Borel set $\widetilde{X} \subset X$, the function $q(\widetilde{X}|x, \cdot)$ is continuous on $A(x)$,

$(S2.4)$ the function $a \mapsto \int_X \omega(y) q(dy|x, a)$ is continuous on $A(x)$ for every $x \in X$.

The above conditions were used in stochastic dynamic programming by many authors, see, e.g., Schäl [30], Bäuerle and Rieder [3], Bertsekas and Shreve [7] or Hernández-Lerma and Lasserre [16,17]. Using the so-called "weight" or "bounding"

function ω one can study dynamic programming models with unbounded one-stage utility u. This method was introduced by Wessels [34], but as noted by van der Wal [32], in the dynamic programming with linear discount function $\delta(z) = \beta z$, one can introduce an extra state $x_e \notin X$, re-define the transition probability and the utility function to obtain an equivalent "bounded model". More precisely, we consider a new state space $X \cup \{x_e\}$, where x_e is an absorbing isolated state. Let $A(x_e) = \{a_e\}$ with an extra action $a_e \notin A$. For $x \in X$ the action sets are $A(x)$. The transition probabilities Q and one-stage utilities R in a new model are as follows

$$Q(B|x,a) := \frac{1}{\alpha\omega(x)} \int_B \omega(y)q(dy|x,a), \quad \text{for } B \in B(X),$$

$$Q(x_e|x,a) := 1 - \frac{\int_X \omega(y)q(dy|x,a)}{\alpha\omega(x)},$$

$$R(x,a) := \frac{u(x,a)}{\omega(x)}, \quad \text{for } (x,a) \in D, \quad R(x_e,a_e) := 0.$$

Here, α is a constant from assumption $(B2.4)$. This transformed Markov decision process is equivalent to the original one in the sense that every policy gives the same total expected discounted payoff up to the factor $\omega(x)$, where $x \in X$ denotes the initial state. We would like to emphasise that in the non-linear discount function case such a transformation to bounded case is not possible. We need to do some extra work.

Remark 2.5 If u is bounded from above by some constant, then we can skip assumptions (A) and it is enough in assumptions (B) to require $(B2.1)$, $(B2.2)$ and $(B2.3)(i)$. In this case, it suffices to put $\alpha = 1$, $\omega(x) = 1$ for all $x \in X$ and it is easily seen that $(B2.3)(ii)$, $(B2.4)$, $(S2.4)$, $(W2.4)$ and $(W2.5)$ hold. If, on the other hand, u is unbounded in the sense that there exists a function ω meeting conditions $(A2.2)$, then $\omega(x) \geq 1$ for all $x \in X$ must be unbounded as well. From Remark 2.3, it follows that $(B2.3)$ holds when the function $z \mapsto \gamma(z)/z$ is non-increasing. We would like to emphasise that condition (ii) in $(B2.3)$ is crucial in our proofs in the case when $(A2.2)$ holds with an unbounded function ω. The dynamic programming problems when only $(A2.2)$ is assumed can be solved by a "truncation method" and then making use of an approximation by solutions for models that satisfy conditions (A).

3 Discounted Utility Evaluations: Two Alternative Approaches

Let $r_1(x,a) = u(x,a)$ for $(x,a) \in D$ and, for any $n \in \mathbb{N}$ and $(h_n, a_n) = (x_1, a_1, \ldots, x_n, a_n) \in D^n$,

$$\begin{aligned}
r_{n+1}&(h_{n+1}, a_{n+1})\\
&= u(x_1,a_1) + \delta\big(u(x_2,a_2) + \delta\big(u(x_3,a_3) + \cdots + \delta\left(u\left(x_{n+1},a_{n+1}\right)\right)\cdots\big)\big)\\
&= u(x_1,a_1) + \delta(r_n(x_2,a_2,\ldots,x_{n+1},a_{n+1})).
\end{aligned}$$

Below in this subsection we assume that all expectations (integrals) and limits exist. In the sequel, we shall study cases where this assumption is satisfied.

Let $\pi \in \Pi$ and $x = x_1 \in X$ be an initial state. By $\mathbb{E}_x^\pi$ we denote the expectation operator with respect to the unique probability measure $\mathbb{P}_x^\pi$ on H induced by the policy $\pi \in \Pi$ and the transition probability q according to the Ionescu–Tulcea theorem, see Proposition 7.28 in [7].

Definition 3.1 For any $\pi = (\pi_k) \in \Pi$ and any initial state $x = x_1$ the n-stage expected discounted utility is

$$R_n(x, \pi) = \mathbb{E}_x^\pi \big[r_n(x_1, a_1, \ldots, x_n, a_n) \big]$$

and the expected discounted utility over an infinite horizon is

$$R(x, \pi) := \lim_{n \to \infty} R_n(x, \pi). \tag{3.1}$$

A policy $\pi^* \in \Pi$ is optimal in the dynamic programming model under utility evaluation (3.1), if

$$R(x, \pi^*) \geq R(x, \pi) \quad \text{for all} \quad \pi \in \Pi, \ x \in X.$$

Remark 3.2 Utility functions as in (3.1) have been considered by Jaśkiewicz et al. [20]. Optimal policies have been shown to exist for the model with *u bounded from above* satisfying assumptions (A) and either (W) or (S). However, optimal policies obtained in [20] are *history-dependent* and are characterised by an infinite system of Bellman equations as in the non-stationary model of Hinderer [18].

To obtain *stationary optimal policies*, we shall define a *recursive discounted utility* using the ideas similar to those developed in papers on dynamic programming by Denardo [12] and Bertsekas [6] and in papers on economic dynamic optimisation [1,2,4,9,14,27,28,33]. The seminal article for these studies was the work by Koopmans [25] on stationary recursive utility generalising the standard discounted utility of Samuelson [29]. To define the recursive discounted utility we must introduce some operator notation.

Let $\pi = (\pi_1, \pi_2, \ldots) \in \Pi$ and $v \in \mathcal{M}(H_{k+1})$. We set

$$Q_{\pi_k}^\delta v(h_k) = \int_X \delta(v(h_k, \pi_k(h_k), x_{k+1})) q(dx_{k+1}|x_k, \pi_k(h_k))$$

and

$$\begin{aligned} T_{\pi_k} v(h_k) &= u(x_k, \pi_k(h_k)) + Q_{\pi_k}^\delta v(h_k) \\ &= u(x_k, \pi_k(h_k)) + \int_X \delta(v(h_k, \pi_k(h_k), x_{k+1})) q(dx_{k+1}|x_k, \pi_k(h_k)). \end{aligned}$$

These operators are well-defined for example when u and v are bounded from above. Similarly, we define $Q_{\pi_k}^\gamma$ with δ replaced by γ. Observe that by $(B2.1)$

$$|Q_{\pi_k}^\delta v(h_k)| \leq Q_{\pi_k}^\gamma |v|(h_k)$$

provided that $Q_{\pi_k}^\delta v(h_k) > -\infty$. This fact will be used frequently.

The interpretation of $T_{\pi_k} v(h_k)$ is as follows. If $x_{k+1} \mapsto v(h_k, \pi_k(h_k), x_{k+1})$ is a "continuation value" of utility, then $T_{\pi_k} v(h_k)$ is the expected discounted utility given the pair $(x_k, \pi_k(h_k))$.

The composition $T_{\pi_1} \circ T_{\pi_2} \circ \cdots \circ T_{\pi_n}$ of the operators $T_{\pi_1}, T_{\pi_2}, \ldots, T_{\pi_n}$ is for convenience denoted by $T_{\pi_1} T_{\pi_2} \cdots T_{\pi_n}$.

Let $\mathbf{0}$ be a function that assigns zero to each argument $y \in X$.

Definition 3.3 For any $\pi = (\pi_k) \in \Pi$ and any initial state $x = x_1$, the n-stage recursive discounted utility is defined as

$$U_n(x, \pi) = T_{\pi_1} \cdots T_{\pi_n} \mathbf{0}(x)$$

and the recursive discounted utility over an infinite horizon is

$$U(x, \pi) := \lim_{n \to \infty} U_n(x, \pi). \tag{3.2}$$

A policy $\pi^* \in \Pi$ is optimal in the dynamic programming model under utility evaluation (3.2), if

$$U(x, \pi^*) \geq U(x, \pi) \quad \text{for all} \quad \pi \in \Pi, \ x \in X. \tag{3.3}$$

For instance, below we give a full formula for $n = 3$. Namely,

$$
\begin{aligned}
U_3(x, \pi) = T_{\pi_1} T_{\pi_2} T_{\pi_3} \mathbf{0}(x) = u(x, \pi_1(x)) + \int_X \delta \Big(u(x_2, \pi_2(x, \pi_1(x), x_2)) \\
+ \int_X \delta(u(x_3, \pi_3(x, \pi_1(x), x_2, \pi_2(x, \pi_1(x), x_2), x_3)) q(dx_3 | x_2, \\
\pi_2(x, \pi_1(x), x_2))) \Big) \\
\times q(dx_2 | x, \pi_1(x)).
\end{aligned}
$$

We would like to point out that in the special case of linear discount function $\delta(z) = \beta z$ with $\beta \in (0, 1)$, the two above-mentioned approaches coincide. In that case we deal with the usual expected discounted utility, because

$$R_n(x, \pi) = U_n(x, \pi) = \mathbb{E}_x^\pi \left[\sum_{k=1}^n \beta^{k-1} u(x_k, a_k) \right].$$

4 Auxiliary Results

Let Y be a Borel space. By $\mathcal{U}(Y)$ we denote the space of all upper semicontinuous functions on Y. We recall some results on measurable selections, a generalisation of the Banach fixed point theorem and present a property of a subadditive function γ.

Lemma 4.1 *Assume that $A(x)$ is compact for each $x \in X$.*

(a) *Let $g \in \mathcal{M}(D)$ be such that $a \mapsto g(x, a)$ is upper semicontinuous on $A(x)$ for each $x \in X$. Then,*

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

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

$$f^*(x) \in \text{Arg} \max_{a \in A(x)} g(x, a)$$

for all $x \in X$.

(b) *If, in addition, we assume that $x \mapsto A(x)$ is upper semicontinuous and $g \in \mathcal{U}(D)$, then $g^* \in \mathcal{U}(X)$.*

Part (a) follows from Corollary 1 in [10]. Part (b) is a corollary to Berge's maximum theorem, see [5, pp. 115–116] and Proposition 10.2 in [30].

Let $\mathcal{M}_b^a(X)$ be the space of all functions $v \in \mathcal{M}(X)$ such that $x \mapsto v(x)/\omega(x)$ is bounded from above on X. The symbol $\mathcal{M}_b^d(X)$ is used for the subspace of functions $v \in \mathcal{M}_b^a(X)$ such that $x \mapsto |v(x)|/\omega(x)$ is bounded on X. Let $\mathcal{M}_b^a(D)$ be the space of all functions $w \in \mathcal{M}(D)$ such that $(x, a) \mapsto w(x, a)/\omega(x)$ is bounded from above on D. By $\mathcal{M}_b^d(D)$ we denote the space of all functions $w \in \mathcal{M}_b^a(D)$ such that $(x, a) \mapsto |w(x, a)|/\omega(x)$ is bounded on D. We also define

$$\mathcal{U}_b^a(X) := \mathcal{M}_b^a(X) \cap \mathcal{U}(X), \quad \mathcal{U}_b^d(X) := \mathcal{M}_b^d(X) \cap \mathcal{U}(X), \quad \text{and}$$
$$\mathcal{U}_b^a(D) := \mathcal{M}_b^a(D) \cap \mathcal{U}(D), \quad \mathcal{U}_b^d(D) := \mathcal{M}_b^d(D) \cap \mathcal{U}(D).$$

Lemma 4.2 *Let assumptions (B) be satisfied and*

$$I(v)(x, a) := \int_X \delta(v(y))q(dy|x, a), \quad v \in \mathcal{M}_b^a(X), \ (x, a) \in D.$$

Then $I(v) \in \mathcal{M}_b^a(D)$. If $v \in \mathcal{M}_b^d(X)$, then $I(v) \in \mathcal{M}_b^d(D)$.

Proof Let $v \in \mathcal{M}_b^a(X)$ and $v^+(y) = \max\{v(y), 0\}$. Then there exists $c_0 > 0$ such that $v(y) \le v^+(y) \le c_0\omega(y)$ for all $y \in X$. Obviously, we have $I(v)(x, a) \le I(v^+)(x, a) \le I(c_0\omega)(x, a) \le \alpha c_0\omega(x)$ for all $(x, a) \in D$. Hence $I(v) \in \mathcal{M}_b^a(D)$. Now assume that $v \in \mathcal{M}_b^d(X)$. Then there exists a constant $c_1 > 0$ such that $|v(y)| \le c_1\omega(y)$ for all $y \in X$ and we obtain $|I(v)(x, a)| \le \alpha c_1\omega(x)$ for all $(x, a) \in D$. Thus $I(v) \in \mathcal{M}_b^d(D)$. □

Our results will be formulated using the standard dynamic programming operators. For any $v \in \mathcal{M}_b^a(X)$, put

$$Sv(x, a) := u(x, a) + \int_X \delta(v(y))q(dy|x, a), \quad (x, a) \in D.$$

Next define

$$Tv(x) := \sup_{a \in A(x)} Sv(x, a) = \sup_{a \in A(x)} \left[u(x, a) + \int_X \delta(v(y))q(dy|x, a) \right]. \quad (4.1)$$

By $T^{(m)}$ we denote the composition of T with itself m times.

If $f \in F$ and $v \in \mathcal{M}_b^a(X)$, then we put

$$T_f v(x) := Sv(x, f(x)) = u(x, f(x)) + \int_X \delta(v(y))q(dy|x, f(x)). \quad (4.2)$$

Clearly, $T_f v \in \mathcal{M}_b^a(X)$.

The next result follows from Lemmas 4.1–4.2 and Lemmas 8.3.7 and 8.5.5 from [17].

Lemma 4.3 *Assume that assumptions (A) and (B) hold.*

(a) *If conditions (W) are also satisfied and $v \in \mathcal{U}_b^d(X)$, then $Sv \in \mathcal{U}_b^d(D)$ and $Tv \in \mathcal{U}_b^d(X)$.*

(b) *If $(S2.2)$–$(S2.4)$ hold and $v \in \mathcal{M}_b^d(X)$, then $Sv \in \mathcal{M}_b^d(D)$ and, for each $x \in X$, the function $a \mapsto Sv(x, a)$ is upper semicontinuous on $A(x)$. Moreover, $Tv \in \mathcal{U}_b^d(X)$.*

Remark 4.4 (a) The assumption that δ is continuous and increasing is important for part (a) of Lemma 4.3.

(b) Under assumptions of Lemma 4.3, in the operator in (4.1) one can replace sup by max.

(c) Using Lemma 4.2, one can easily see that if $v \in \mathcal{M}_b^a(X)$ and $f \in F$, then $T_f v \in \mathcal{M}_b^a(X)$.

The following fixed point theorem will play an important role in our proof (see e.g. [26] or Theorem 5.2 in [13]).

Lemma 4.5 *Let (Z, m) be a complete metric space, $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous, increasing function with $\psi(x) < x$ for all $x \in (0, \infty)$. If an operator $T : Z \to Z$ satisfies the inequality*

$$m(Tv_1, Tv_2) \le \psi(m(v_1, v_2))$$

for all $v_1, v_2 \in Z$, then T has a unique fixed point $v^* \in Z$ and

$$\lim_{n \to \infty} m(T^{(n)}v, v^*) = 0$$

for each $v \in Z$. Here $T^{(n)}$ is the composition of T with itself n times.

For the convenience of the reader we formulate and prove a modification of Lemma 8 from [20] that is used many times in our proofs. Consider a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ and put

$$\psi_m(z) = z + \psi\big(z + \psi\big(z + \cdots + \psi(z + \psi(z))\cdots\big)\big), \quad \text{where } z > 0 \text{ appears } m \text{ times.}$$

Lemma 4.6 *If ψ is increasing, subadditive and $\psi(y) < y$ for all $y > 0$, then for any $z > 0$, there exists*

$$L(z) := \lim_{m \to \infty} \psi_m(z) = \sup_{m \geq 1} \psi_m(z) < \infty.$$

Proof For any $k \in \mathbb{N}$, let $\psi^{(k)}$ mean the composition of ψ with itself k times. Note that since the function ψ is increasing, then for each $m \geq 1$,

$$\psi_{m+1}(z) > \psi_m(z).$$

Hence, the sequence $(\psi_m(z))$ is increasing. We show that its limit is finite. Indeed, observe that by the subadditivity of ψ, we have

$$\psi_2(z) - \psi_1(z) = z + \psi(z) - z \leq \psi(z), \quad \text{and}$$
$$\psi_3(z) - \psi_2(z) = z + \psi(z + \psi(z)) - z - \psi(z) \leq \psi^{(2)}(z).$$

By induction, we obtain

$$\psi_m(z) - \psi_{m-1}(z) \leq \psi^{(m-1)}(z).$$

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

$$\psi_m(z) - \psi_{m-1}(z) < \epsilon - \psi(\epsilon).$$

Observe now that from subadditivity of ψ (set $\gamma := \psi$ in (2.1)), it follows that

$$
\begin{aligned}
\psi_{m+1}(z) - \psi_{m-1}(z) &= \psi_{m+1}(z) - \psi_m(z) + \psi_m(z) - \psi_{m-1}(z) \\
&\leq z + \psi(\psi_m(z)) - z - \psi(\psi_{m-1}(z)) + \epsilon - \psi(\epsilon) \\
&\leq \psi(\psi_m(z) - \psi_{m-1}(z)) + \epsilon - \psi(\epsilon) \\
&\leq \psi(\epsilon - \psi(\epsilon)) + \epsilon - \psi(\epsilon) < \psi(\epsilon) + \epsilon - \psi(\epsilon) = \epsilon.
\end{aligned}
$$

By induction, we can easily prove that

$$\psi_{m+k}(z) - \psi_{m-1}(z) \leq \epsilon$$

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

5 Stationary Optimal Policies in Dynamic Problems with the Recursive Discounted Utilities

In this section, we prove that if $u \in \mathcal{M}_b^a(D)$, assumptions (B) hold and either conditions (W) or (S) are satisfied, then the recursive discounted utility functions (3.2) are

well-defined and there exists an optimal stationary policy. Moreover, under assumptions (W) $((S))$, the value function $x \mapsto \sup_{\pi \in \Pi} U(x, \pi)$ belongs to $\mathcal{U}_b^a(X)$ $(\mathcal{M}_b^a(X))$. The value function and an optimal policy will be characterised via a single Bellman equation. First we shall study the case $u \in \mathcal{M}_b^d(D)$ and then apply an approximation technique to the unbounded from below case.

5.1 One-Period Utilities with Bounds on Both Sides

Assume that $\mathcal{M}_b^d(X)$ is endowed with the so-called weighted norm $\| \cdot \|_\omega$ defined as

$$\|v\|_\omega := \sup_{x \in X} \frac{|v(x)|}{\omega(x)}, \quad v \in \mathcal{M}_b^d(X).$$

Then $\mathcal{M}_b^d(X)$ is a Banach space and $\mathcal{U}_b^d(X)$ is a closed subset of $\mathcal{M}_b^d(X)$, if $(W2.4)$ holds. The following theorem is the main result of this subsection. Its proof is split in different parts below.

Theorem 5.1 *Suppose that assumptions* (A), (B) *hold and assumptions* (W) *are satisfied. Then*

(a) *the Bellman equation* $Tv = v$ *has a unique solution* $v^* \in \mathcal{M}_b^d(X)$ *and*

$$\lim_{m \to \infty} \|T^{(m)} \mathbf{0} - v^*\|_\omega = 0 \quad and \quad v^*(x) = \sup_{\pi \in \Pi} U(x, \pi), \ x \in X,$$

(b) *there exists* $f^* \in F$ *such that* $T_{f^*} v^* = v^*$ *and* f^* *is an optimal stationary policy for problem* (3.3),
(c) $v^* \in \mathcal{U}_b^d(X)$.
The points (a) and (b) also remain valid under assumptions (A), (B) *and* (S).

Remark 5.2 As already mentioned, in our approach we consider only deterministic strategies. This is because the optimality results do not change, when we take randomised strategies into account. Actually, we may examine a new model in which the original action sets $A(x)$ are replaced by the set of probability measures $\Pr(A(x))$. Then, the Bellman equation has a solution as in Theorem 5.1, but the supremum in (4.1) is taken over the set $\Pr(A(x))$. However, due to our assumptions the maximum is also attained at a Dirac delta concentrated at some point from $A(x)$. Therefore, randomised strategies do not influence the results.

Since condition $(B2.4)$ contains two cases, it is convenient to define a new function

$$\tilde{\gamma}(y) := \begin{cases} \gamma(y), & \alpha \leq 1 \\ \alpha \gamma(y), & \alpha > 1. \end{cases}$$

Clearly, $\tilde{\gamma}$ is subadditive. Let $z = \max\{b, c\}$, the constants $b > 0$ and $c > 0$ come from (A). Then $|u(x, a)| \leq \omega(x) z$ for all $(x, a) \in D$. From $(B2.3)(ii)$, it follows

that

$$\tilde{\gamma}(\omega(x)y) \le \omega(x)\tilde{\gamma}(y), \quad \text{for all} \quad x \in X, \ y \ge 0. \tag{5.1}$$

This inequality is frequently used in our proofs. Let

$$\tilde{\gamma}_k(z) = z + \tilde{\gamma}\big(z + \tilde{\gamma}(z + \cdots + \tilde{\gamma}(z))\cdots\big),$$

where z appears on the right-hand side k times. Putting $\psi = \tilde{\gamma}$ in Lemma 4.6, we infer that

$$\tilde{L}(z) := \lim_{k \to \infty} \tilde{\gamma}_k(z) = \sup_{k \in \mathbb{N}} \tilde{\gamma}_k(z) < \infty.$$

$$\tag{5.2}$$

We point out that $\tilde{\gamma}^{(n)}$ is the n-th iteration of the function $\tilde{\gamma}$.

We now prove that the recursive discounted utility (3.2) is well-defined.

Lemma 5.3 *If* $u \in \mathcal{M}_b^d(D)$ *and assumptions* (B) *are satisfied, then* $U(x, \pi) := \lim_{n \to \infty} U_n(x, \pi)$ *exists for any policy* $\pi \in \Pi$ *and any initial state* $x \in X$. *Moreover,* $U(\cdot, \pi) \in \mathcal{M}_b^d(X)$ *and*

$$\lim_{n \to \infty} \|U(\cdot, \pi) - U_n(\cdot, \pi)\|_\omega = 0.$$

Proof We shall prove that $(U_n(\cdot, \pi))$ is a Cauchy sequence of functions in $\mathcal{M}_b^d(X)$ for each policy $\pi \in \Pi$. We claim that

$$|U_{n+m}(x, \pi) - U_n(x, \pi)| \le Q_{\pi_1}^\gamma \cdots Q_{\pi_{n-1}}^\gamma Q_{\pi_n}^\gamma |T_{\pi_{n+1}} \cdots T_{\pi_{n+m}} \mathbf{0}|(x). \tag{5.3}$$

Indeed, using assumptions (B), we can conclude that

$$\begin{aligned}
|U_{n+m}(x, \pi) - U_n(x, \pi)| &= |T_{\pi_1} \cdots T_{\pi_n} T_{\pi_{n+1}} \cdots T_{\pi_{n+m}} \mathbf{0}(x) - T_{\pi_1} \cdots T_{\pi_n} \mathbf{0}(x)| \\
&= |Q_{\pi_1}^\delta T_{\pi_2} \cdots T_{\pi_{n+m}} \mathbf{0}(x) - Q_{\pi_1}^\delta T_{\pi_2} \cdots T_{\pi_n} \mathbf{0}(x)| \\
&\le Q_{\pi_1}^\gamma |T_{\pi_2} \cdots T_{\pi_{n+m}} \mathbf{0} - T_{\pi_2} \cdots T_{\pi_n} \mathbf{0}|(x) \le \quad (\text{cont.})\cdots \\
&\le Q_{\pi_1}^\gamma \cdots Q_{\pi_{n-1}}^\gamma |Q_{\pi_n}^\delta T_{\pi_{n+1}} \cdots T_{\pi_{n+m}} \mathbf{0}|(x) \\
&\le Q_{\pi_1}^\gamma \cdots Q_{\pi_{n-1}}^\gamma Q_{\pi_n}^\gamma |T_{\pi_{n+1}} \cdots T_{\pi_{n+m}} \mathbf{0}|(x).
\end{aligned}$$

Assume that $m = 1$. Then, for any $h_{n+1} \in H_{n+1}$, we have

$$|T_{\pi_{n+1}} \mathbf{0}(h_{n+1})| = |u(x_{n+1}, \pi_{n+1}(h_{n+1}))| \le \omega(x_{n+1})z.$$

Take $m = 2$ and notice by (B) that

$$
\begin{aligned}
|T_{\pi_{n+1}} T_{\pi_{n+2}} \mathbf{0}(h_{n+1})| &= |u(x_{n+1}, \pi_{n+1}(h_{n+1})) + \int_X \delta(u(x_{n+2}, \pi_{n+2}(h_{n+1}, x_{n+2}))) \\
&\qquad q(dx_{n+2}|x_{n+1}, \pi_{n+1}(h_{n+1}))| \\
&\leq \omega(x_{n+1})z + \int_X \gamma(\omega(x_{n+2})z)q(dx_{n+2}|x_{n+1}, \pi_{n+1}(h_{n+1})) \\
&\leq \omega(x_{n+1})z + \int_X \omega(x_{n+2})\gamma(z)q(dx_{n+2}|x_{n+1}, \pi_{n+1}(h_{n+1})) \\
&\leq \omega(x_{n+1})(z + \alpha\gamma(z)) \leq \omega(x_{n+1})(z + \tilde{\gamma}(z)).
\end{aligned}
$$

For $m = 3$, it follows that

$$
\begin{aligned}
|T_{\pi_{n+1}} T_{\pi_{n+2}} T_{\pi_{n+3}} \mathbf{0}(h_{n+1})| &= |u(x_{n+1}, \pi_{n+1}(h_{n+1})) \\
&\quad + \int_X \delta(T_{\pi_{n+2}} T_{\pi_{n+3}} \mathbf{0}(h_{n+1}, \pi_{n+1}(h_{n+1}), x_{n+2}))q(dx_{n+2}|x_{n+1}, \pi_{n+1}(h_{n+1}))| \\
&\leq \omega(x_{n+1})z + \int_X \gamma\left(\omega(x_{n+2})(z + \tilde{\gamma}(z))\right) q(dx_{n+2}|_{n+1}, \pi_{n+1}(h_{n+1})) \\
&\leq \omega(x_{n+1})z + \int_X \omega(x_{n+2})\gamma\left(z + \tilde{\gamma}(z)\right) q(dx_{n+2}|x_{n+1}, \pi_{n+1}(h_{n+1})) \\
&\leq \omega(x_{n+1})z + \alpha\omega(x_{n+1})\gamma(z + \tilde{\gamma}(z)) \leq \omega(x_{n+1})(z + \tilde{\gamma}(z + \tilde{\gamma}(z))).
\end{aligned}
$$

Continuing this way, for any $h_{n+1} \in H_{n+1}$, we obtain

$$
\begin{aligned}
|T_{\pi_{n+1}} \cdots T_{\pi_{n+m}} \mathbf{0}(h_{n+1})| &\leq \omega(x_{n+1})\left(z + \tilde{\gamma}\left(z + \tilde{\gamma}(z + \cdots + \tilde{\gamma}(z + \tilde{\gamma}(z))\cdots)\right)\right) \\
&= \omega(x_{n+1})\tilde{\gamma}_m(z),
\end{aligned} \tag{5.4}
$$

where z appears m times on the right-hand side of inequality (5.4). By (5.2), $\tilde{\gamma}_m(z) < \tilde{L}(z) < \infty$. Combining (5.3) and (5.4) and making use of $(B2.4)$ and (5.1), we conclude that

$$
\begin{aligned}
Q_{\pi_1}^\gamma \cdots Q_{\pi_{n-1}}^\gamma Q_{\pi_n}^\gamma \tilde{L}(z)\omega(x) &\leq Q_{\pi_1}^\gamma \cdots Q_{\pi_{n-1}}^\gamma \gamma(\tilde{L}(z))\alpha\omega(x) \\
&= Q_{\pi_1}^\gamma \cdots Q_{\pi_{n-1}}^\gamma \tilde{\gamma}(\tilde{L}(z))\omega(x) \cdots \\
&\leq \tilde{\gamma}^{(n)}(\tilde{L}(z))\omega(x).
\end{aligned}
$$

Consequently,

$$
\|U_{n+m}(\cdot, \pi) - U_n(\cdot, \pi)\|_\omega \leq \tilde{\gamma}^{(n)}(\tilde{L}(z)). \tag{5.5}
$$

From the proof of (5.4) we deduce that for any $n \in \mathbb{N}$

$$
|U_n(x, \pi)| \leq \omega(x)\tilde{\gamma}_n(z) \leq \omega(x)\tilde{L}(z).
$$

Therefore, for each $n \in \mathbb{N}$, $U_n(\cdot, \pi) \in \mathcal{M}_b^d(X)$. From (5.5) it follows that $(U_n(x, \pi))$ is a Cauchy sequence in the Banach space $\mathcal{M}_b^d(X)$. $\qquad\square$

Proof of Theorem 5.1 Consider first assumptions (W). By Lemma 4.3, T maps $\mathcal{U}_b^d(X)$ into itself. We show that T has a fixed point in $\mathcal{U}_b^d(X)$. Let $v_1, \ v_2 \in \mathcal{U}_b^d(X)$. Then, under assumptions (B) we obtain

$$
\begin{aligned}
|Tv_1(x) - Tv_2(x)| &\leq \sup_{a \in A(x)} \int_X \left| \delta\big(v_1(y)\big) - \delta\big(v_2(y)\big) \right| q(dy|x, a) \\
&\leq \sup_{a \in A(x)} \int_X \gamma\big(|v_1(y) - v_2(y)|\big) q(dy|x, a) \\
&\leq \sup_{a \in A(x)} \left| \int_X \gamma\big(\|v_1 - v_2\|\big)\omega(y) q(dy|x, a) \right| \\
&\leq \alpha\gamma\big(\|v_1 - v_2\|_\omega\big)\omega(x).
\end{aligned}
$$

Hence,

$$
\|Tv_1 - Tv_2\|_\omega \leq \tilde{\gamma}(\|v_1 - v_2\|_\omega).
$$

Since the space $\mathcal{U}_b^d(X)$ endowed with the metric induced by the norm $\|\cdot\|_\omega$ is complete, by Lemma 4.5, there exists a unique $v^* \in \mathcal{U}_b^d(X)$ such that $v^* = Tv^*$ and

$$
\lim_{n \to \infty} \|T^{(n)}v - v^*\|_\omega = 0 \quad \text{for any} \quad v \in \mathcal{U}_b^d(X).
$$

By Lemma 4.1 and the assumptions that δ is increasing and continuous, it follows that there exists $f^* \in F$ such that $v^* = T_{f^*}v^*$. We claim that

$$
v^*(x) = U(x, f^*) = \lim_{n \to \infty} T_{f^*}^{(n)} v^*(x) \quad \text{for all } x \in X.
$$

The operator $T_{f^*} : \mathcal{M}_b^d(X) \to \mathcal{M}_b^d(X)$ also satisfies assumptions of Lemma 4.5. Thus there is a unique function $\tilde{v} \in \mathcal{M}_b^d(X)$ such that

$$
\tilde{v}(x) = T_{f^*}\tilde{v}(x) = \lim_{n \to \infty} T_{f^*}^{(n)} h(x), \quad x \in X,
$$

for any $h \in \mathcal{M}_b^d(X)$. Therefore, $\tilde{v} = v^*$. Putting $h := \mathbf{0}$ we deduce from Lemma 5.3 that

$$
\lim_{n \to \infty} T_{f^*}^{(n)} \mathbf{0}(x) = U(x, f^*), \quad x \in X.
$$

In order to prove the optimality of f^* note that for any $a \in A(x)$ and $x \in X$, it holds

$$
v^*(x) \geq u(x, a) + \int_X \delta(v^*(y)) q(dy|x, a).
$$

Taking any policy $\pi = (\pi_n)$ and iterating the above inequality, we get

$$v^*(x) \geq T_{\pi_1} \cdots T_{\pi_n} v^*(x), \quad x \in X.$$

We now prove that

$$\lim_{n \to \infty} T_{\pi_1} \cdots T_{\pi_n} v^*(x) = U(x, \pi), \quad x \in X.$$

With this end in view, we first consider the differences

$$|U_n(x, \pi) - T_{\pi_1} \cdots T_{\pi_n} v^*(x)| = |T_{\pi_1} \cdots T_{\pi_n} \mathbf{0}(x) - T_{\pi_1} \cdots T_{\pi_n} v^*(x)|$$
$$\leq \omega(x) \tilde{\gamma}^{(n)}(\|v^*\|_\omega) \to 0 \quad \text{as } n \to \infty.$$

By Lemma 5.3, $U_n(x, \pi) \to U(x, \pi)$ for every $x \in X$ as $n \to \infty$. Therefore, we have that

$$v^*(x) \geq \lim_{n \to \infty} T_{\pi_1} \cdots T_{\pi_n} v^*(x) = U(x, \pi), \quad x \in X$$

and

$$\sup_{\pi \in \Pi} U(x, \pi) \geq U(x, f^*) = v^*(x) \geq \sup_{\pi \in \Pi} U(x, \pi), \quad x \in X.$$

This implies that

$$U(x, f^*) = v^*(x) = \sup_{\pi \in \Pi} U(x, \pi), \quad x \in X,$$

which finishes the proof under assumptions (W). For assumptions (S) the proof proceeds along the same lines. By Lemma 4.3, under (S), $T : \mathcal{M}_b^d(X) \to \mathcal{M}_b^d(X)$. $\qquad \square$

Remark 5.4 Under assumptions of Theorem 5.1, the Bellman equation has a unique solution and it is the optimal value function $v^*(x) = \sup_{\pi \in \Pi} U(x, \pi)$. Moreover, it holds that

$$\lim_{n \to \infty} \|T^{(n)} \mathbf{0} - v^*\|_\omega = 0.$$

Obviously, $T^{(n)} \mathbf{0}$ is the value function in the n-step dynamic programming problem. One can say that the *value iteration algorithm* works and the iterations $T^{(n)} \mathbf{0}(x)$ converge to $v^*(x)$ for each $x \in X$. This convergence is uniform in $x \in X$ when the weight function ω is bounded.

5.2 One-Period Utilities Unbounded from Below

In this subsection we drop condition $(A2.1)$ and assume that there exists $c > 0$ such that $u(x, a) \leq c\omega(x)$ for all $(x, a) \in D$. In other words, $u \in \mathcal{M}_b^a(D)$. Here we obtain the following result which is shown in the remaining part of this subsection.

Theorem 5.5 *Suppose that assumptions* $(A2.2)$, (B) *and* (W) *are satisfied. Then*

(a) *the optimal value function*

$$\underline{v}^*(x) := \sup_{\pi \in \Pi} U(x, \pi), \ x \in X,$$

   is a solution to the Bellman equation $Tv = v$ *and* $\underline{v}^* \in \mathcal{M}_b^a(X)$,
(b) *there exists* $\tilde{f} \in F$ *such that* $T_{\tilde{f}}\underline{v}^* = \underline{v}^*$ *and* $\tilde{f}$ *is an optimal stationary policy,*
(c) $\underline{v}^* \in \mathcal{U}_b^a(X)$.
   The points (a) and (b) also remain valid under assumptions $(A2.2)$, (B) *and* (S).

Remark 5.6 We shall prove that $\underline{v}^*$ is the limit of a non-increasing sequence of value functions in "truncated models", i.e. the models that satisfy $(A2.1)$ and $(A2.2)$. The convergence is monotone, but it is not uniform. The Bellman equation may have many unbounded solutions.

Remark 5.7 The assumptions of Theorem 5.5 do not guarantee uniqueness. An example is very simple. Assume that $X = \mathbb{N}$, $A = A(x) = \{a\}$, $u(x, a) = 0$ for all $(x, a) \in D$, and the process moves from state x to $x + 1$ with probability one. The discount function $\delta(x) = \beta x$ with $\beta \in (0, 1)$. Clearly, u satisfies assumption $(A2.2)$ with $\omega(x) = 1$, $c = 1$. Note that $v(x) = r/\beta^x$ is a solution to the Bellman equation $Tv = v$ for any $r \in \mathbb{R}$. Clearly, $\underline{v}^*(x) = 0$ is one of them. Actually, $\underline{v}^*(x) = 0$ is the largest non-positive solution to the Bellman equation. This example does not contradict the uniqueness result in Theorem 5.1. Within the class of bounded functions $\underline{v}^*(x) = 0$ is the unique solution to the Bellman equation.

We now prove that the recursive discounted utility (3.2) is well-defined.

Lemma 5.8 *If* $u \in \mathcal{M}_b^a(D)$ *and assumptions* (B) *are satisfied, then* $U(x, \pi) := \lim_{n \to \infty} U_n(x, \pi)$ *exists in* $\mathbb{R}$ *for any policy* $\pi \in \Pi$ *and any initial state* $x \in X$. *Moreover,* $U(\cdot, \pi) \in \mathcal{M}_b^a(X)$.

Our assumption that $u \in \mathcal{M}_b^a(D)$ means that $(A2.2)$ holds, i.e., there exists $c > 0$ such that $u(x, a) \leq c\omega(x)$ for all $(x, a) \in D$.

Proof of Lemma 5.8 We divide the proof into five parts.
Step 1: We start with a simple observation: for any $n \in \mathbb{N}$, $x \in X$ and $\pi \in \Pi$ it holds

$$U_{n+1}(x, \pi) \leq U_n(x, \pi) + \tilde{\gamma}^{(n)}(c)\omega(x).$$

From assumptions (B), it follows that

$$\delta(a + c\omega(x)) \leq \delta(a) + \gamma(c\omega(x)) \leq \gamma(c)\omega(x)$$

for $a \in \mathbb{R}$ and $x \in X$. Note that, for any $h_k \in H_k$ and π_k, we have

$$
\begin{aligned}
T_{\pi_k} c\omega(h_k) &= u(x_k, \pi_k(h_k)) + \int_X \delta(c\omega(y)) q(dy | x_k, \pi_k(h_k)) \\
&\leq u(x_k, \pi_k(h_k)) + Q^{\gamma}_{\pi_k} c\omega(x_k) \leq T_{\pi_k} \mathbf{0}(h_k) + \tilde{\gamma}(c)\omega(x_k). \quad (5.6)
\end{aligned}
$$

Furthermore, for any $v \in \mathcal{M}(H_{k+1})$ such that $v(h_k, \pi_k(h_k), y) \leq \eta\omega(y)$ for all $y \in X$ and some $\eta > 0$, we obtain

$$
\begin{aligned}
&T_{\pi_k}(v + c\omega)(h_k) \\
&= u(x_k, \pi_k(h_k)) + \int_X \delta(v(h_k, \pi_k(h_k), y) + c\omega(y)) q(dy | x_k, \pi_k(h_k)) \\
&\leq u(x_k, \pi_k(h_k)) + Q^{\delta}_{\pi_k} v(h_k) + Q^{\gamma}_{\pi_k} c\omega(h_k) \leq T_{\pi_k} v(h_k) + \tilde{\gamma}(c)\omega(x_k).
\end{aligned}
$$

From this fact and (5.6) we conclude that

$$
\begin{aligned}
U_{n+1}(x, \pi) = T_{\pi_1} \cdots T_{\pi_n} T_{\pi_{n+1}} \mathbf{0}(x) &\leq T_{\pi_1} \cdots T_{\pi_{n-1}} T_{\pi_n} c\omega(x) \\
&\leq T_{\pi_1} \cdots T_{\pi_{n-1}} \big(T_{\pi_n} \mathbf{0} + \tilde{\gamma}(c)\omega\big)(x) \\
&\leq T_{\pi_1} \cdots T_{\pi_{n-2}} \big(T_{\pi_{n-1}} T_{\pi_n} \mathbf{0} + \tilde{\gamma}^{(2)}(c)\omega\big)(x) \ldots \text{ (cont.)} \\
&\leq T_{\pi_1} \cdots T_{\pi_{n-2}} T_{\pi_{n-1}} T_{\pi_n} \mathbf{0}(x) + \tilde{\gamma}^{(n)}(c)\omega(x) \\
&= U_n(x, \pi) + \tilde{\gamma}^{(n)}(c)\omega(x). \quad (5.7)
\end{aligned}
$$

This finishes the first step.

Step 2: Let $U_n(x, \pi) = -\infty$ for some $n \in \mathbb{N}$, $x \in X$ and $\pi \in \Pi$. Then, by Step 1, $U_m(x, \pi) = -\infty$ for all $m \geq n$. Therefore, $\lim_{m \to \infty} U_m(x, \pi) = -\infty$.

Step 3: Let $v \in \mathcal{M}(H_{k+1})$ be such that $v(h_k, a_k, x_{k+1}) \leq \eta\omega(x_{k+1})$ for every $x_{k+1} \in X$ and for some $\eta > 0$. Define

$$
\begin{aligned}
\Gamma_{\pi_k} v(h_k) &:= c\omega(x_k) + Q^{\gamma}_{\pi_k} v(h_k) \quad \text{and} \\
\Gamma^{\pi_{n+m}}_{\pi_{n+1}} \mathbf{0}(h_{n+1}) &:= \Gamma_{\pi_{n+1}} \cdots \Gamma_{\pi_{n+m}} \mathbf{0}(h_{n+1}).
\end{aligned}
$$

Note that

$$
\Gamma_{\pi_{n+m}} \mathbf{0}(h_{n+m}) = c\omega(x_{n+m}).
$$

Next, we have

$$
\begin{aligned}
&\Gamma_{\pi_{n+m-1}} \Gamma_{\pi_{n+m}} \mathbf{0}(h_{n+m-1}) \\
&= c\omega(x_{n+m-1}) + \int_X \gamma\big(c\omega(x_{n+m})\big) q\big(dx_{n+m} | x_{n+m-1}, \pi_{n+m-1}(h_{n+m-1})\big) \\
&\leq \omega(x_{n+m-1})(c + \tilde{\gamma}(c)),
\end{aligned}
$$

and

$$\Gamma_{\pi_{n+m-2}}\Gamma_{\pi_{n+m-1}}\Gamma_{\pi_{n+m}}\mathbf{0}(h_{n+m-2})$$
$$\leq c\omega(x_{n+m-2}) + \int_X \gamma\big(\omega(x_{n+m-1})(c + \tilde{\gamma}(c))\big)$$
$$q\big(dx_{n+m-1}|x_{n+m-2}, \pi_{n+m-2}(h_{n+m-2})\big)$$
$$\leq \omega(x_{n+m-2})(c + \tilde{\gamma}(c + \tilde{\gamma}(c))).$$

Continuing this way, we get

$$\Gamma_{\pi_{n+1}}^{\pi_{n+m}}\mathbf{0}(h_{n+1}) = \Gamma_{\pi_{n+1}} \cdots \Gamma_{\pi_{n+m}}\mathbf{0}(h_{n+1})$$
$$\leq \omega(x_{n+1})\big(c + \tilde{\gamma}\big(c + \tilde{\gamma}\big(c + \cdots + \tilde{\gamma}(c + \tilde{\gamma}(c))\cdots\big)\big)\big),$$

where c appears on the right-hand side of this inequality m times. Putting $\tilde{c} = \tilde{L}(z)$ with $z = c$ in (5.2), we obtain

$$\Gamma_{\pi_{n+1}}^{\pi_{n+m}}\mathbf{0}(h_{n+1}) = \Gamma_{\pi_{n+1}} \cdots \Gamma_{\pi_{n+m}}\mathbf{0}(h_{n+1}) \leq \omega(x_{n+1})\tilde{c} = \omega(x_{n+1})\tilde{L}(c) < \infty.$$
$$\tag{5.8}$$

Step 4: For $m, n \in \mathbb{N}$, we set

$$W_{n,m}(x, \pi) := T_{\pi_1} \cdots T_{\pi_n}\Gamma_{\pi_{n+1}}^{\pi_{n+m}}\mathbf{0}(x) \quad \text{and}$$
$$W_{n,0}(x, \pi) := T_{\pi_1} \cdots T_{\pi_n}\mathbf{0}(x) = U_n(x, \pi).$$

For any $k = 1, \ldots, n-1$ and $h_{k+1} \in H_{k+1}, \pi \in \Pi$, let $\pi(k+1) = (\pi_{k+1}, \pi_{k+2}, \ldots)$ and

$$V_{n-k,m}^{\pi(k+1)}(h_{k+1}) := T_{\pi_{k+1}} \cdots T_{\pi_n}\Gamma_{\pi_{n+1}}^{\pi_{n+m}}\mathbf{0}(h_{k+1}),$$
$$V_{n-k,0}^{\pi(k+1)}(h_{k+1}) := T_{\pi_{k+1}} \cdots T_{\pi_n}\mathbf{0}(h_{k+1}).$$

For $k = n - 1$, we have

$$V_{1,m}^{\pi(n)}(h_n) = T_{\pi_n}\Gamma_{\pi_{n+1}}^{\pi_{n+m}}\mathbf{0}(h_n),$$

and

$$V_{1,0}^{\pi(n)}(h_n) = T_{\pi_n}\mathbf{0}(h_n) = u(x_n, \pi_n(h_n)).$$

Hence and from (5.8), it follows that

$$V_{1,m}^{\pi(n)}(h_n) - V_{1,0}^{\pi(n)}(h_n) \leq \int_X \gamma\big(\omega(x_{n+1})\tilde{c}\big)q(dx_{n+1}|x_n, \pi_n(h_n)) \leq \omega(x_n)\tilde{\gamma}(\tilde{c}).$$
$$\tag{5.9}$$

Observe that for each $k = 1, \ldots, n-2$,

$$
\begin{aligned}
V_{n-k,m}^{\pi(k+1)}(h_{k+1}) - V_{n-k,0}^{\pi(k+1)}(h_{k+1}) &= T_{\pi_{k+1}} V_{n-k-1,m}^{\pi(k+2)}(h_{k+1}) - T_{\pi_{k+1}} V_{n-k-1,0}^{\pi(k+2)}(h_{k+1}) \\
&= Q_{\pi_{k+1}}^{\delta} V_{n-k-1,m}^{\pi(k+2)}(h_{k+1}) - Q_{\pi_{k+1}}^{\delta} V_{n-k-1,0}^{\pi(k+2)}(h_{k+1}) \\
&\leq Q_{\pi_{k+1}}^{\gamma} \big(V_{n-k-1,m}^{\pi(k+2)} - V_{n-k-1,0}^{\pi(k+2)} \big)(h_{k+1}).
\end{aligned} \tag{5.10}
$$

It is important to note that

$$
V_{n-k,m}^{\pi(k+1)} - V_{n-k,0}^{\pi(k+1)} > 0, \quad k = 1, \ldots, n-1.
$$

Now, using (5.10), for any $\pi \in \Pi$ and $x = x_1$, we conclude that

$$
\begin{aligned}
W_{n,m}(x, \pi) - W_{n,0}(x, \pi) &= T_{\pi_1} V_{n-1,m}^{\pi(2)}(x) - T_{\pi_1} V_{n-1,0}^{\pi(2)}(x) \\
&= Q_{\pi_1}^{\delta} V_{n-1,m}^{\pi(2)}(x) - Q_{\pi_1}^{\delta} V_{n-1,0}^{\pi(2)}(x) \\
&\leq Q_{\pi_1}^{\gamma} \big(V_{n-1,m}^{\pi(2)} - V_{n-1,0}^{\pi(2)} \big)(x) \\
&\leq Q_{\pi_1}^{\gamma} Q_{\pi_2}^{\gamma} \big(V_{n-2,m}^{\pi(3)} - V_{n-2,0}^{\pi(3)} \big)(x) \ldots \text{ (cont.)} \\
&\leq Q_{\pi_1}^{\gamma} Q_{\pi_2}^{\gamma} \cdots Q_{\pi_{n-1}}^{\gamma} \big(V_{1,m}^{\pi(n)} - V_{1,0}^{\pi(n)} \big)(x).
\end{aligned}
$$

This and (5.9) imply that

$$
W_{n,m}(x, \pi) - W_{n,0}(x, \pi) \leq \omega(x) \tilde{\gamma}^{(n)}(\tilde{c}), \quad \text{for all } m, \, n \in \mathbb{N}. \tag{5.11}
$$

Step 5: We now consider the case where $U_n(x, \pi) > -\infty$ for an initial state $x \in X$, a policy $\pi \in \Pi$ and for all $n \in \mathbb{N}$. From (5.8) we have

$$
W_{n,m}(x, \pi) \leq W_{n,m+1}(x, \pi) \leq \Gamma_{\pi_1}^{\pi_{n+m+1}} \mathbf{0}(x) \leq \tilde{c}\omega(x) < \infty.
$$

Therefore, $\lim_{m \to \infty} W_{n,m}(x, \pi)$ exists and is finite. Let us denote this limit by G_n. Note that, for each $m, \, n \in \mathbb{N}$,

$$
U_n(x, \pi) = W_{n,0}(x, \pi) \leq W_{n,m}(x, \pi).
$$

Let $\epsilon > 0$ be fixed. Then, by (5.11), for sufficiently large n, say $n > N_0$,

$$
W_{n,m}(x, \pi) \leq W_{n,0}(x, \pi) + \epsilon
$$

for all $m \in \mathbb{N}$. Thus

$$
W_{n,m}(x, \pi) - \epsilon \leq W_{n,0}(x, \pi) \leq W_{n,m}(x, \pi)
$$

and consequently

$$
G_n - \epsilon \leq W_{n,0}(x, \pi) \leq G_n
$$

for all $n > N_0$. Observe that the sequence (G_n) is non-increasing and $G_* := \lim_{n \to \infty} G_n$ exists in the extended real line $\overline{\mathbb{R}}$. Hence, the limit

$$\lim_{n \to \infty} W_{n,0}(x, \pi) = \lim_{n \to \infty} U_n(x, \pi)$$

also exists and equals G_*. □

In the proof of Theorem 5.5 we shall need the following result (see [30] or Theorem A.1.5 in [3]).

Lemma 5.9 *If Y is a metric space and (w_n) is a non-increasing sequence of upper semicontinuous functions $w_n : Y \to \overline{\mathbb{R}}$, then*

(a) $w_\infty = \lim_{n \to \infty} w_n$ *exists and w_∞ is upper semicontinuous,*
(b) *if, additionally, Y is compact, then*

$$\max_{y \in Y} \lim_{n \to \infty} w_n(y) = \lim_{n \to \infty} \max_{y \in Y} w_n(y).$$

In the proof of Theorem 5.5 we shall refer to the dynamic programming operators defined in (4.1) and (4.2). Moreover, we also define corresponding operators for $v \in \mathcal{M}_b^a(X)$ and $K \in \mathbb{N}$ as follows

$$T^K v(x) := \sup_{a \in A(x)} \left[u^K(x, a) + \int_X \delta(v(y)) q(dy | x, a) \right], \quad x \in X,$$

and

$$T_f^K v(x) = u^K(x, f(x)) + \int_X \delta(v(y)) q(dy | x, f(x)), \quad x \in X,$$

where $f \in F$ and $u^K(x, a) = \max\{u(x, a), 1 - K\}$, $K \in \mathbb{N}$. The recursive discounted utility functions with one-period utility u^K in the finite (n-periods) and infinite time horizon for an initial state $x \in X$ and a policy $\pi \in \Pi$ will be denoted by $U_n^K(x, \pi)$ and $U^K(x, \pi)$, respectively.

Lemma 5.10 *For any $n \in \mathbb{N}$ and $f \in F$, it holds*

$$\lim_{K \to \infty} U_n^K(x, f) = U_n(x, f), \quad x \in X.$$

Proof We proceed by induction. For $n = 1$ the fact is obvious. Suppose that

$$T_f^{K,(n)} \mathbf{0}(x) = U_n^K(x, f) \to T_f^{(n)} \mathbf{0}(x) = U_n(x, f) \quad \text{as } K \to \infty, \text{ for all } x \in X.$$

Here, $T_f^{K,(n)}$ denotes the n-th composition of the operator T_f^K with itself. Then, by our induction hypothesis, our assumption that δ is continuous and increasing, and the

monotone convergence theorem, we infer that, for every $x \in X$,

$$\lim_{K \to \infty} T_f^K T_f^{K,(n)} \mathbf{0}(x) = \lim_{K \to \infty} \left(u^K(x, f(x)) + \int_X \delta(T_f^{K,(n)} \mathbf{0}(y)) q(dy|x, f(x)) \right)$$

$$= u(x, f(x)) + \int_X \delta(T_f^{(n)} \mathbf{0}(y)) q(dy|x, f) = T_f^{(n+1)} \mathbf{0}(x).$$

The lemma now follows by the induction principle.   □

Proof of Theorem 5.5 Assume first (W). The proof for (S) is analogous with obvious changes. By Theorem 5.1, for any $K \in \mathbb{N}$, there exists a unique solution $v^{*,K} \in \mathcal{U}_b^d(X)$ to the Bellman equation and, for each $x \in X$, $v^{*,K}(x) = \sup_{\pi \in \Pi} U^K(x, \pi)$. Since, $u^K \geq u$, it follows that

$$v^{*,K}(x) = \sup_{\pi \in \Pi} U^K(x, \pi) \geq \sup_{\pi \in \Pi} U(x, \pi) = \underline{v}^*(x), \quad x \in X.$$

Clearly, the sequence $(v^{*,K})$ is non-increasing, thus $v_\infty(x) := \lim_{K \to \infty} v^{*,K}(x)$ exists in $\mathbb{R}$ for every $x \in X$, and consequently,

$$v_\infty(x) \geq \underline{v}^*(x), \quad x \in X. \tag{5.12}$$

From Theorem 5.1 we know that $v^{*,K}$ is a solution to the equation

$$v^{*,K}(x) = \sup_{a \in A(x)} \left[u^K(x, a) + \int_X \delta(v^{*,K}(y)) q(dy|x, a) \right], \quad x \in X.$$

Since both sequences $(v^{*,K}(x))$, $x \in X$, and $(u^K(x, a))$, $(x, a) \in D$, are non-increasing, it follows from Lemma 5.9, our assumption that δ is increasing and continuous and the monotone convergence theorem that

$$v_\infty(x) = \lim_{K \to \infty} v^{*,K}(x) = \lim_{K \to \infty} \max_{a \in A(x)} \left[u^K(x, a) + \int_X \delta(v^{*,K}(y)) q(dy|x, a) \right]$$

$$= \max_{a \in A(x)} \lim_{K \to \infty} \left[u^K(x, a) + \int_X \delta(v^{*,K}(y)) q(dy|x, a) \right]$$

$$= \max_{a \in A(x)} \left[u(x, a) + \int_X \delta(v_\infty(y)) q(dy|x, a) \right], \quad x \in X. \tag{5.13}$$

Moreover, in case (W), we have that $v_\infty \in \mathcal{U}_b^a(X)$. From the obvious inequalities $u(x, a) \leq u^1(x, a) \leq c\omega(x)$, $(x, a) \in D$, it follows that $v_\infty(x) \leq \tilde{c}\omega(x)$ for $\tilde{c} = \tilde{L}(c)$ and for all $x \in X$ (put $z = c$ in (5.2)). By Lemma 4.1, there exists a maximiser $\tilde{f} \in F$ on the right-hand side of equation (5.13) and we have

$$v_\infty(x) = u(x, \tilde{f}(x)) + \int_X \delta(v_\infty(y)) q(dy|x, \tilde{f}(x)) = T_{\tilde{f}} v_\infty(x), \quad x \in X.$$

Iterating this equation, we obtain that

$$v_\infty(x) = T_{\tilde{f}}^{(n)} v_\infty(x) \le T_{\tilde{f}}^{K,(n)} v_\infty(x) \le T_{\tilde{f}}^{K,(n)} \tilde{c}\omega(x), \text{ for all } x \in X \text{ and } k \in \mathbb{N}.$$

From (5.7) in the proof of Lemma 5.8, (with c replaced by $\tilde{c}$, u replaced by u^K and $\pi_1 = \cdots = \pi_n = \tilde{f}$) we infer that

$$v_\infty(x) \le T_{\tilde{f}}^{K,(n)} \tilde{c}\omega(x) \le U_n^K(x, \tilde{f}) + \tilde{\gamma}^{(n)}(\tilde{c})\omega(x), \text{ for all } x \in X \text{ and } n \in \mathbb{N}.$$

Letting $K \to \infty$ in the above inequality and making use of Lemma 5.10 yield that

$$v_\infty(x) \le U_n(x, \tilde{f}) + \tilde{\gamma}^{(n)}(\tilde{c})\omega(x) \text{ for all } x \in X.$$

Hence,

$$v_\infty(x) \le \lim_{n\to\infty} \left(U_n(x, \tilde{f}) + \tilde{\gamma}^{(n)}(\tilde{c})\omega(x) \right) = U(x, \tilde{f}) \le \sup_{\pi \in \Pi} U(x, \pi)$$
$$= \underline{v}^*(x), \quad x \in X.$$

From this inequality and (5.12), we conclude that

$$v_\infty(x) = U(x, \tilde{f}) = \sup_{\pi \in \Pi} U(x, \pi) = \underline{v}^*(x), \text{ for all } x \in X,$$

and the proof is finished.　　　　　　　　　　　　　　　　　　　　　　□

6 Computational Issues

In this section we consider the unbounded utility setting as in Theorem 5.1.

6.1 Policy Iteration

An optimal stationary policy can be computed as a limit point of a sequence of decision rules. In what follows, we define $V_0 = \mathbf{0}$ and $V_n := T^{(n)}\mathbf{0}$ for $n \in \mathbb{N}$. Next for fixed $x \in X$, let

$$A_n^*(x) := \text{Arg} \max_{a \in A(x)} \left(u(x, a) + \int_X \delta(V_{n-1}(y))q(dy|x, a) \right)$$

for $n \in \mathbb{N}$. In the same way, let

$$A^*(x) := \text{Arg} \max_{a \in A(x)} \left(u(x, a) + \int_X \delta(v^*(y))q(dy|x, a) \right).$$

By $Ls\,A_n^*(x)$, we denote the *upper limit of the set sequence* $(A_n^*(x))$, that is, the set of all *accumulation points* of sequences (a_n) with $a_n \in A_n^*(x)$ for all $n \in \mathbb{N}$.

The next result states that an optimal stationary policy can be obtained from accumulation points of sequences of maximisers of recursively computed value functions. Related results for dynamic programming with standard discounting are discussed, for example, in [3,30].

Theorem 6.1 *Under assumptions of Theorem* 5.1*, we obtain:* $\emptyset \neq Ls\,A_n^*(x) \subset A^*(x)$ *for all* $x \in X$.

Proof Fix $x \in X$ and define for $n \in \mathbb{N}$ the functions $v_n : A(x) \to \mathbb{R}$ by

$$v_n(a) := u(x, a) + \int_X \delta(V_{n-1}(y))q(dy|x, a).$$

By Lemma 4.3, v_n is upper semicontinuous on $A(x)$. Moreover, for $m \geq n$ and $a \in A(x)$, we have

$$|v_m(a) - v_n(a)| \leq |\int_X \delta\big(T^{(m-1)}\mathbf{0}(y)\big)q(dy|x, a) - \int_X \delta\big(T^{(n-1)}\mathbf{0}(y)\big)q(dy|x, a)|$$

$$\leq \int_X \gamma\big(|T^{(m-1)}\mathbf{0}(y) - T^{(n-1)}\mathbf{0}(y)|\big)q(dy|x, a). \tag{6.1}$$

Using the arguments as in the proof of (5.5), we infer that

$$|T^{(m-1)}\mathbf{0}(y) - T^{(n-1)}\mathbf{0}(y)| \leq \sup_{\pi \in \Pi} |U_{m-1}(y, \pi) - U_{n-1}(y, \pi)|$$

$$\leq \omega(y)\sup_{\pi \in \Pi} \|U_{m-1}(\cdot, \pi) - U_{n-1}(\cdot, \pi)\|_\omega \leq \omega(y)\tilde{\gamma}^{(n-1)}(\tilde{L}(z)). \tag{6.2}$$

From (6.1) and (6.2), it follows that

$$|v_m(a) - v_n(a)| \leq \int_X \gamma\big(\omega(y)\tilde{\gamma}^{(n-1)}(\tilde{L}(z))\big)q(dy|x, a)$$

$$\leq \int_X \omega(y)\gamma\big(\tilde{\gamma}^{(n-1)}(\tilde{L}(z))\big)q(dy|x, a) \leq \omega(x)\tilde{\gamma}^{(n)}\big(\tilde{L}(z)\big).$$

for all $a \in A(x)$. Hence

$$\max_{a \in A(x)} |v_m(a) - v_n(a)| \leq \omega(x)\tilde{\gamma}^{(n)}\big(\tilde{L}(z)\big) =: \varepsilon_n.$$

This then implies that $v_m(a) \leq v_n(a) + \varepsilon_n$ for $m \geq n$ and all $a \in A(x)$. Since $\varepsilon_n \to 0$ as $n \to \infty$, the result follows from Theorem A.1.5. in [3] and the fact that $v^* = \lim_{n\to\infty} V_n$ (see Remark 5.4). $\qquad\square$

6.2 Howard's Policy Improvement Algorithm

The algorithm proposed by Howard [19] is widely discussed in the literature on Markov decision processes (dynamic programming), see [3,16,17] and their references. It may be also applied to models with the recursive discounted utility.

For any $f \in F$ we shall use the following notation $U_f = U(\cdot, f)$.

Theorem 6.2 *Let assumptions of Theorem* 5.1 *be satisfied. For any* $f \in F$, *denote*

$$A(x, f) := \{a \in A(x) \mid u(x, a) + \int_X \delta(U(y, f))q(dy|x, a) > U(x, f)\}, \quad x \in X.$$

Then, the following holds:

(a) *If for some Borel set* $X_0 \subset X$ *we define a decision rule* g *by*

$$
\begin{aligned}
g(x) &\in A(x, f) \quad \text{for } x \in X_0, \\
g(x) &= f(x) \quad \text{for } x \notin X_0,
\end{aligned}
$$

then $U_g \geq U_f$ *and* $U_g(x) > U_f(x)$ *for* $x \in X_0$. *In this case, the policy* g *is called* an improvement *of* f.

(b) *If* $A(x, f) = \emptyset$ *for all* $x \in X$, *then* $U_f = v^*$, *i.e., the stationary policy* $f \in F$ *is optimal.*

Proof (a) From the definition of g we obtain

$$T_g U_f(x) > U(x, f),$$

if $x \in X_0$ and $T_g U_f(x) = U(x, f)$, if $x \notin X_0$. Thus by induction

$$U_f(x) \leq T_g U_f(x) \leq T_g^{(n)} U_f(x),$$

where the first inequality is strict for $x \in X_0$. Letting $n \to \infty$, it follows as in the proof of Theorem 5.1 that $U_f \leq U_g$ and in particular $U(x, f) < U(x, g)$ for $x \in X_0$.

(b) The condition $A(x, f) = \emptyset$ for all $x \in X$ implies $T U_f \leq U_f$. Since we always have $T U_f \geq T_f U_f = U_f$, we obtain $T U_f = U_f$. From Theorem 5.1 we know that T has a unique fixed point $v^* \in \mathcal{U}_b^d(X)$ (under assumptions (W)) or $v^* \in \mathcal{M}_b^d(X)$ (under assumptions (S)). Thus $U_f = v^*$. □

Altogether, we have the following algorithm for the computation of the value function and an optimal stationary policy:

1. Choose $f_0 \in F$ arbitrary and set $k = 0$.
2. Compute U_{f_k} as the unique solution $v \in \mathcal{M}_b^d(X)$ of the equation $v = T_{f_k} v$.
3. Choose $f_{k+1} \in F$ such that

$$f_{k+1}(x) \in \text{Arg} \max_{a \in A(x)} \left(u(x, a) + \int_X \delta(U_{f_k}(y))q(dy|x, f_k(x)) \right)$$

and set $f_{k+1}(x) = f_k(x)$ if possible. If $f_{k+1} = f_k$, then $U_{f_k} = v^*$ and $(f_k, f_k, \ldots)$ is an optimal stationary policy. Else set $k := k + 1$ and go to step 2.

It is obvious that the algorithm stops in a finite number of steps if the state and action sets are finite. In general, as in the standard discounted case (see, e.g., Theorem 7.5.1 and Corollary 7.5.3 in [3]) we can only claim that

$$v^*(x) = \lim_{k \to \infty} U(x, f_k).$$

7 Applications

Example 7.1 *(Stochastic optimal growth model 1)* There is a single good available to the consumer. The level of this good at the beginning of period $t \in \mathbb{N}$ is given by $x_t \in X := [0, \infty)$. The consumer has to divide x_t between consumption $a_t \in A := [0, \infty)$ and investment (saving) $y_t = x_t - a_t$. Thus $A(x_t) = [0, x_t]$. From consumption a_t the consumer receives utility $u(x_t, a_t) = \sqrt{a_t}$. Investment, on the other hand, is used for production with input y_t yielding output

$$x_{t+1} = y_t \cdot \xi_t, \qquad t \in \mathbb{N},$$

where (ξ_t) is a sequence of i.i.d. shocks with distribution ν being a probability measure on $[0, \infty)$. The initial state $x = x_1 \in \mathbb{R}_+$ is non-random. Further, we assume that

$$\bar{s} = \mathbb{E}\xi_t = \int_0^\infty s\nu(ds) \leq 1.$$

Let the discount function be as follows

$$\delta(z) = (1 - \varepsilon)z + \varepsilon \ln(1 + z), \quad z \geq 0$$

with some $\varepsilon \in (0, 1)$. We observe that there is no constant $\beta \in (0, 1)$ such that $\delta(z) < \beta z$ for all $z > 0$. We define $\gamma := \delta$ and note that $z \mapsto \gamma(z)/z$ is non-increasing. Hence, γ and δ satisfy assumptions $(B2.1)$-$(B2.3)$. Now we show that assumptions (A), (W) and $(B2.4)$ are satisfied with an appropriate function ω. With this end in view, put

$$\omega(x) = \sqrt{x + 1}, \quad x \in X.$$

Then $|u(x, a)| \leq \sqrt{x} \leq \sqrt{x + 1}$ for $a \in A(x) = [0, x]$ and $x \in X$. Thus, (A) holds. Furthermore, by Jensen's inequality we have

$$\int_X \omega(y)q(dy|x, a) = \int_0^\infty \sqrt{(x - a)s + 1}\, \nu(ds) \leq \sqrt{(x - a)\bar{s} + 1} \leq \sqrt{x + 1}$$

for $a \in A(x)$ and $x \in X$. Hence, $(B2.4)$ is satisfied with $\alpha = 1$. It is obvious that conditions (W) are also met. Therefore, from Theorem 5.1, there exists an upper

semicontinuous solution to the Bellman equation and a stationary optimal policy $f^* \in F$.

Example 7.2 *(Stochastic optimal growth model 2)* We study a modified model from Example 1. We assume that the next state evolves according to the equation

$$x_{t+1} = y_t^\theta \cdot \xi_t + (1 - \rho) \cdot y_t, \qquad t \in \mathbb{N}$$

where ρ, θ are some constants from the interval $(0, 1)$. Here, (ξ_t) is again a sequence of i.i.d. shocks with distribution ν and the expected value $\bar{s} > 0$. The utility function is $u(x, a) = a^\sigma$ with $\sigma \in (0, 1)$. Let $\omega(x) = (x + r)^\sigma$, where $r \geq 1$. Assume now that the discount function is of the form

$$\delta(z) = (1 - 2\varepsilon)z + \varepsilon \ln(1 + z), \quad z \geq 0$$

with some $\varepsilon \in (0, 1)$. Then, $\delta(z) \leq (1 - \varepsilon)z$ for $z \geq 0$. By Example 1 in [2] (see also Example 2 in [23]), we have that

$$\int_X \omega(y)q(dy|x, a) \leq \left(1 + \frac{(\bar{s}/\rho)^{\frac{1}{1-\theta}}}{r}\right)^\sigma \omega(x), \quad x \in X.$$

Hence, in $(B2.4)$, we set

$$\alpha := \left(1 + \frac{(\bar{s}/\rho)^{\frac{1}{1-\theta}}}{r}\right)^\sigma > 1$$

and $\alpha\gamma(z) < z$ is satisfied with $\gamma := \delta$ and r such that $\alpha(1 - \varepsilon) < 1$. Clearly, all conditions (A), (B) and (W) are satisfied. $(B(2.3)$ follows from the fact that $z \mapsto \gamma(z)/z$ is non-increasing.) By Theorem 5.1, the value function v^* is upper semicontinuous and satisfies the Bellman equation. Moreover, there exists an optimal stationary policy.

Example 7.3 *(Inventory model)* A manager inspects the stock at each period $t \in \mathbb{N}$. The number of units in stock is $x_t \in X := [0, \infty)$. He can sell $\min\{x_t, \Delta_t\}$ units in period t, where $\Delta_t \geq 0$ is a random variable representing an unknown demand. At the end of period t he can also order any amount $a_t \in A := [0, \hat{a}]$ of new goods to be delivered at the beginning of next period at a cost $C(a_t)$ paid immediately. Here $\hat{a}$ is some positive constant. Moreover, the function C is bounded above by $\hat{C}$, lower semicontinuous, increasing and $C(0) = 0$. The state equation is of the form

$$x_{t+1} = x_t - \min\{x_t, \Delta_t\} + a_t, \quad \text{for } t \in \mathbb{N},$$

where (Δ_t) is a sequence of i.i.d. random variables such that each Δ_t follows a continuous distribution Φ and $\mathbb{E}\Delta_t < \infty$. The manager considers a recursive discounted utility with a discount function δ satisfying $(B2.1)$–$(B2.3)$ (with $\omega = 1$). This model

can be viewed as a Markov decision process, in which $u(x, a) := p\mathbb{E} \min\{x, \Delta\} - C(a)$ is the one period bounded utility function and p denotes the unit stock price. (Here $\Delta = \Delta_t$ for any fixed $t \in \mathbb{N}$.) Clearly, $-\hat{C} \leq u(x, a) \leq p\mathbb{E}\Delta$. Next note that the transition probability q is of the form

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

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

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

depends continuously on (x, a). Hence, the model satisfies assumptions $(W2.1)$-$(W2.3)$. Therefore, by Theorem 5.1, there exists a bounded upper semicontinuous solution to the Bellman equation and an optimal stationary policy $f^* \in F$.

Example 7.4 *(Stopping problem)* We now describe a stopping problem with non-linear discounting. Suppose the Borel state space is X and there is an (uncontrolled) Markov chain with an initial distribution q_0 and the transition probability $q(\cdot|x)$. By $\mathbb{P}$ we denote the probability measure on the product space X^∞ of all trajectories of the Markov chain induced by q_0 and q. At each period the controller has to decide whether to stop the process and receive the reward $R(x)$, where x is the current state or to continue. In the latter case the reward $C(x)$ (which might be a negative cost) is received. The aim is to find a stopping time such that the recursive discounted reward is maximized. We assume here that the controller has to stop with probability one. This problem is a special case of the more general model of Sect. 2. We have to choose here

   (i) $X \cup \{\infty\}$ is the state space where ∞ is an absorbing state which indicates that the process is already stopped,
  (ii) $A := \{0, 1\}$ where $a = 0$ means continue and $a = 1$ means stop,
 (iii) $A(x) := A$ for all $x \in X \cup \{\infty\}$,
 (iv) $q(B|x, 0) := q(B|x)$ for $x \in X$, B a Borel set and $q(\{\infty\}|x, 1) = 1$ for $x \in X$ and $q(\infty|\infty, \cdot) = 1$,
  (v) $u(x, a) := C(x)(1 - a) + R(x)a$ for $x \in X$ and $u(\infty, \cdot) = 0$.

We assume now that $|C(x)| \leq \omega(x)$ and $|R(x)| \leq \omega(x)$ which implies (A) and we assume (B). The optimisation problem is considered with the recursive discounted utility and an interpretation that the receiving of rewards (costs) is stopped after a random time. This random time is a stopping time with respect to the filtration generated by observable states (for more details see Chapter 10 in [3]).

Proposition 7.5 *If the above assumptions on the stopping problem are satisfied, then*

(a) *there exists a function $v^* \in \mathcal{M}_b^d(X)$ such that*

$$v^*(x) = \max \left\{ R(x); \ C(x) + \int_X \delta(v^*(y)) q(dy|x) \right\}. \qquad (7.1)$$

(b) *Moreover, define $f^*(x) = 1$ if $v^*(x) = R(x)$ and $f^*(x) = 0$ else and let $\tau^* := \inf\{n \in \mathbb{N} : f^*(x_n) = 1\}$. If $\mathbb{P}(\tau^* < \infty) = 1$, then τ^* is optimal for the stopping problem and $v^*(x) = \sup_{\pi \in \Pi} U(x, \pi)$.*

The action space A consists of two elements only, so assumptions (S) are satisfied. In the above description we already assumed (A) and (B). Therefore the result follows from Theorem 5.1

Let us consider a special example. Imagine a person who wants to sell her house. At the beginning of each week she receives an offer, which is randomly distributed over the interval $[m, M]$ with $0 < m < M$. The offers are independent and identically distributed with distribution q. The house seller has to decide immediately whether to accept or reject this offer. If she rejects, the offer is lost and she has maintenance cost. Which offer should she accept in order to maximise her expected reward?

Here, we have $X := [m, M]$, $C(x) \equiv -c$ and $R(x) := x$. From Proposition 7.5 we obtain that the value function satisfies

$$v^*(x) = \max \left\{ x; \ -c + \int_{[m,M]} \delta(v_*(y)) q(dy) \right\}.$$

Note that $C^* := \int_{[m,M]} \delta(v^*(y)) q(dy)$ is obviously a constant independent of x. Thus the optimal strategy is to accept the first offer which is above $-c + C^*$. The corresponding stopping time is geometrically distributed and thus certainly satisfies $\mathbb{P}(\tau^* < \infty) = 1$. Moreover, it is not difficult to see that whenever we have two discount functions δ_1, δ_2, which satisfy assumptions (B) and which are ordered, i.e., $\delta_1 \leq \delta_2$, then $C_1^* \leq C_2^*$ because the operator T is monotone. Thus, with stricter discounting we will stop earlier.

References

1. Balbus, Ł.: On recursive utilities with non-affine aggregator and conditional certainty equivalent. Econ. Theory **70**, 551–577 (2019)

2. Bäuerle, N., Jaśkiewicz, A.: Stochastic optimal growth model with risk-sensitive preferences. J. Econ. Theory **173**, 181–200 (2018)
3. Bäuerle, N., Rieder, U.: Markov Decision Processes with Applications to Finance. Springer, Berlin (2011)
4. Becker, R.A., Boyd III, J.H.: Capital Theory. Equilibrium Analysis and Recursive Utility. Blackwell Publishers, New York (1997)
5. Berge, C.: Topological Spaces. MacMillan, New York (1963)
6. Bertsekas, D.P.: Monotone mappings with application in dynamic programming. SIAM J. Control Optim. **15**, 438–464 (1977)
7. Bertsekas, D.P., Shreve, S.E.: Stochastic Optimal Control: The Discrete Time Case. Academic Press, New York (1978)
8. Blackwell, D.: Discounted dynamic programming. Ann. Math. Stat. **36**, 226–235 (1965)
9. Boyd III, J.H.: Recursive utility and the Ramsey problem. J. Econ. Theory **50**, 326–345 (1990)
10. Brown, L.D., Purves, R.: Measurable selections of extrema. Ann. Stat. **1**, 902–912 (1973)
11. Carbon Pricing Dashboard. The World Bank. https://carbonpricingdashboard.worldbank.org/map_data
12. Denardo, E.V.: Contraction mappings in the theory underlying dynamic programming. SIAM Rev. **9**, 165–177 (1967)
13. Dugundji, J., Granas, A.: Fixed Point Theory. Springer, New York (2003)
14. Durán, J.: Discounting long run average growth in stochastic dynamic programs. Econ. Theory **22**, 395–413 (2003)
15. Feinberg, E.A., Shwartz, A. (eds.): Handbook of Markov Decision Processes: Theory and Methods. Kluwer Academic Publishers, Dordrecht (2002)
16. Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York (1996)
17. Hernández-Lerma, O., Lasserre, J.B.: Further Topics on Discrete-Time Markov Control Processes. Springer, New York (1999)
18. Hinderer, K.: Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter Lecture Notes in Operations Research. Springer, New York (1970)
19. Howard, R.A.: Dynamic Programming and Markov Processes. The Technology Press of MIT, Cambridge, MA (1960)
20. Jaśkiewicz, A., Nowak, A.S.: Stochastic games with unbounded payoffs: applications to robust control in economics. Dyn. Games Appl. **1**, 253–279 (2011)
21. Jaśkiewicz, A., Matkowski, J., Nowak, A.S.: Persistently optimal policies in stochastic dynamic programming with generalized discounting. Math. Oper. Res. **38**, 108–121 (2013)
22. Jaśkiewicz, A., Matkowski, J., Nowak, A.S.: On variable discounting in dynamic programming: applications to resource extraction and other economic models. Ann. Oper. Res. **220**, 263–278 (2014)
23. Jaśkiewicz, A., Matkowski, J., Nowak, A.S.: Generalised discounting in dynamic programming with unbounded returns. Oper. Res. Lett. **42**, 231–233 (2014)
24. Kechris, A.S.: Classical Descriptive Set Theory. Springer, New York (1995)
25. Koopmans, T.C.: Stationary ordinal utility and impatience. Econometrica **28**, 287–309 (1960)
26. Matkowski, J.: Integral solutions of functional equations. Dissertationes Math. **127**, 1–68 (1975)
27. Miao, J.: Economic Dynamics in Discrete Time. MIT Press, Cambridge (2014)
28. Ozaki, H., Streufert, P.A.: Dynamic programming for non-additive stochastic objectives. J. Math. Econ. **25**, 391–442 (1996)
29. Samuelson, P.: A note on measurement of utility. Rev. Econ. Stud. **4**, 155–161 (1937)
30. Schäl, M.: Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. Z. Wahrscheinlichkeitstheorie Verwand. Gebiete **32**, 179–196 (1975)
31. Stokey, N.L., Lucas Jr., R.E., Prescott, E.: Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge, MA (1989)
32. van der Wal, J.: Stochastic Dynamic Programming. Mathematical Centre Tracts, Amsterdam (1981)
33. Weil, P.: Nonexpected utility in macroeconomics. Quart. J. Econ. **105**, 29–42 (1990)
34. Wessels, J.: Markov programming by successive approximations with respect to weighted supremum norms. J. Math. Anal. Appl. **58**, 326–335 (1977)