Appl Math Optim (2010) 61: 421–434
DOI 10.1007/s00245-009-9093-x

Two Characterizations of Optimality in Dynamic Programming

Ioannis Karatzas · William D. Sudderth

Published online: 19 November 2009

Abstract It holds in great generality that a plan is optimal for a dynamic programming problem, if and only if it is "thrifty" and "equalizing." An alternative characterization of an optimal plan, that applies in many economic models, is that the plan must satisfy an appropriate Euler equation and a transversality condition. Here we explore the connections between these two characterizations.

Keywords Dynamic programming · Optimality · Thriftiness · Equalization · Euler equations · Envelope equation · Transversality condition

1 Introduction

It was shown by Dubins and Savage [4] that necessary and sufficient conditions for a strategy to be optimal for a gambling problem are that the strategy be "thrifty" and "equalizing." These conditions were later adapted for dynamic programming by

Research supported in part by the National Science Foundation under Grant NSF-DMS-06-01774 and NSF-DMS-09-05754. The first author acknowledges the hospitality of the Mathematics and Statistics Departments of the University of Minnesota, during a visit in May 2008.

I. Karatzas (✉)
Department of Mathematics, Columbia University, MailCode 4438, 2990 Broadway, New York, NY 10027, USA
e-mail: ik@math.columbia.edu

I. Karatzas
INTECH Investment Management LLC, One Palmer Square, Suite 441, Princeton, NJ 08542, USA
e-mail: ik@enhanced.com

W.D. Sudderth
School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA
e-mail: bill@stat.umn.edu

Blackwell [3], Hordijk [6] and Rieder [9], among others. For a special class of dynamic programming problems important in economic models, it has been shown that optimality is equivalent to the satisfaction of an "Euler equation" and a "transversality condition"; see Stokey & Lucas [12] for a discussion and references. Our main objective here is to understand the relationship between these two characterizations of optimality. One corollary of our approach is a simple proof for the necessity of the transversality condition, which has been considered a difficult problem. (See Stokey & Lucas, page 102, and Kamihigashi [7].) The notions of "thrifty" and "equalizing" seem not to be very widely known to dynamic programmers working in economics, though they have proved quite useful in other contexts; we hope that this paper will help spread the word about them.

Section 2 is a brief exposition of the thrifty-and-equalizing theory for a fairly general class of dynamic programming models. Section 3 introduces the Euler equation and the transversality condition, and then explains their relationship to the thrifty and equalizing conditions. In Sect. 4 we take a brief look at "envelope inequalities" and "Euler inequalities" for one-dimensional problems without imposing smoothness or interiority conditions, and obtain the necessity of an appropriate "transversality condition" in this context. There is an Appendix on measure-theoretic questions that arise in dynamic programming.

2 Thrifty and Equalizing

Consider a dynamic programming problem (S, A, q, r, β) where S is a nonempty set of *states*, the mapping A assigns to each state $s \in S$ a nonempty set $A(s)$ of *actions* available at s, the *law of motion* q associates to each pair (s, a) with $s \in S$, $a \in A(s)$ a probability distribution $q(\cdot|s, a)$ on S, the *daily reward* $r(\cdot, \cdot)$ is a nonnegative function defined on pairs (s, a) with $s \in S$ and $a \in A(s)$, and $\beta \in (0, 1)$ is a discount factor. Play begins in some state $s = s_1$; you choose an action $a_1 \in A(s_1)$, receive a reward of $r(s_1, a_1)$, and the system moves to the next state s_2 which is an S-valued random variable with distribution $q(\cdot|s_1, a_1)$. This process is iterated, yielding a random sequence

$$(s_1, a_1),\ (s_2, a_2), \ldots \tag{2.1}$$

and a total discounted reward $\sum_{n=1}^{\infty} \beta^{n-1} r(s_n, a_n)$. A *plan* is a sequence $\pi = (\pi_1, \pi_2, \ldots)$, where π_n tells you how to choose the nth action a_n as a function $\pi_n(h_n)$ of the previous history $h_n = (s_1, a_1, \ldots, s_{n-1}, a_{n-1}, s_n)$. A plan π, together with an initial state $s_1 = s$, determine the distribution $\mathbb{P}^{\pi,s}$ of the random sequence in (2.1) as well as the expected total discounted reward

$$R^{\pi}(s) := \mathbb{E}^{\pi,s}\left(\sum_{n=1}^{\infty} \beta^{n-1} r(s_n, a_n)\right). \tag{2.2}$$

The *optimal reward* or *value* at $s \in S$ is $V(s) := \sup_{\pi} R^{\pi}(s)$.

Remark 1 (Measure theory) If the state space S is uncountable, then nontrivial measure-theoretic questions arise in the theory. For example, it can happen that the value function $V(\cdot)$ is not Borel measurable, even when all the primitives of the problem are Borel in an appropriate sense [2]. To ease the exposition, we defer further discussion of these difficulties to the Appendix, where it will be explained how they can be resolved. For now we ask the reader to suspend disbelief and assume that the functions which arise in this section are measurable and the expectations are well-defined.

Assume throughout, for simplicity, that $V(s) < \infty, \forall s \in S$, and ask whether the supremum in (2.2) is attained, say by some plan π^*; if so, this plan is called *optimal*. A key tool for answering such questions is the characterization of the value function $V(\cdot)$ via the *Bellman equation*

$$V(s) = \sup_{a \in A(s)} \left(r(s,a) + \beta \int_S V(t)\, q(dt|s,a) \right), \quad s \in S$$

which holds in great generality and is also known as the "optimality equation" (see, for example, Sect. 9.4 of Bertsekas & Shreve [1]). For $a \in A(s)$ and a measurable function $\mathfrak{v} : S \mapsto \mathbb{R}^+$, define

$$(T_a \mathfrak{v})(s) := r(s,a) + \beta \int_S \mathfrak{v}(t)\, q(dt|s,a).$$

The Bellman equation can now be cast as: $V(s) = \sup_{a \in A(s)}[(T_a V)(s)]$.

Definition 1 An action $a \in A(s)$ *conserves* $V(\cdot)$ *at* $s \in S$, if $(T_a V)(s) = V(s)$.
Thus an action $a \in A(s)$ conserves $V(\cdot)$ at $s \in S$, if and only if

$$a \in \arg\max_{A(s)} \left\{ r(s,\cdot) + \beta \int_S V(t)\, q(dt|s,\cdot) \right\}.$$

Notice also that $(T_a V)(s) \le V(s)$ for all s and $a \in A(s)$.

Let us fix now an initial state $s = s_1$ along with a plan π, and consider the random sequences $\{M_n\}_{n \ge 1}$ and $\{Q_n\}_{n \ge 1}$ given by

$$Q_n := \sum_{k=1}^{n} \beta^{k-1} r(s_k, a_k), \tag{2.3}$$

$$M_1 := V(s_1), \qquad M_{n+1} := Q_n + \beta^n V(s_{n+1}), \quad n \ge 1. \tag{2.4}$$

Let $\mathcal{F}_n$ be the σ-field generated by the history $h_n = (s_1, a_1, \ldots, s_{n-1}, a_{n-1}, s_n)$.

Lemma 1 *For every plan π and initial state s, the adapted sequences $\{M_n, \mathcal{F}_n\}_{n \ge 1}$ and $\{\beta^{n-1} V(s_n), \mathcal{F}_n\}_{n \ge 1}$ are nonnegative supermartingales under $\mathbb{P}^{\pi,s}$.*

Proof Set $Q_0 = 0$. Then for any $n \ge 1$, any given history $h_n = (s_1, a_1, \ldots, s_{n-1}, a_{n-1}, s_n)$, and letting $a_n = \pi_n(h_n)$, we have

$$M_{n+1} = Q_{n-1} + \beta^{n-1}\big[r(s_n, a_n) + \beta V(s_{n+1})\big], \quad \text{and thus}$$

$$\mathbb{E}^{\pi,s}[M_{n+1}|\mathcal{F}_n] = Q_{n-1} + \beta^{n-1}\,(T_{a_n}V)(s_n) \le Q_{n-1} + \beta^{n-1}\,V(s_n) = M_n \tag{2.5}$$

a.s. under $\mathbb{P}^{\pi,s}$; we have used the fact that Q_n is $\mathcal{F}_n$-measurable, and $q(\cdot\,|s_n, a_n)$ is the conditional distribution of s_{n+1} given $\mathcal{F}_n$. Thus $\{M_n, \mathcal{F}_n\}_{n\ge 1}$ is a $\mathbb{P}^{\pi,s}$-supermartingale. The sequence $\{Q_n\}_{n\ge 1}$ is nondecreasing, since the daily reward function $r(\cdot,\cdot)$ is nonnegative. From this fact and (2.4), it follows easily that $\{\beta^{n-1}V(s_n), \mathcal{F}_n\}_{n\ge 1}$ is also a $\mathbb{P}^{\pi,s}$-supermartingale. All of these sequences are clearly nonnegative, because $r(\cdot\,,\cdot)$ is. □

It follows from the lemma that the sequences $\{M_n\}_{n\ge 1}$ and $\{\beta^{n-1}V(s_n)\}_{n\ge 1}$ converge almost surely and are non-increasing in expectation, under $\mathbb{P}^{\pi,s}$. Define

$$\Lambda^\pi(s) := \lim_{n\to\infty} \downarrow \mathbb{E}^{\pi,s}(M_n),$$

so that

$$\begin{aligned} V(s) = \mathbb{E}^{\pi,s}(M_1) &\ge \lim_{n\to\infty} \mathbb{E}^{\pi,s}(M_{n+1}) = \Lambda^\pi(s) \\ &= \lim_{n\to\infty} \left\{\mathbb{E}^{\pi,s}(Q_n) + \beta^n\,\mathbb{E}^{\pi,s}[V(s_{n+1})]\right\} \\ &= R^\pi(s) + \lim_{n\to\infty} \left\{\beta^n\,\mathbb{E}^{\pi,s}[V(s_{n+1})]\right\} \ge R^\pi(s). \end{aligned} \tag{2.6}$$

Definition 2 A given plan π is called: *thrifty* at $s \in S$, if $V(s) = \Lambda^\pi(s)$; it is called *equalizing* at $s \in S$, if $\Lambda^\pi(s) = R^\pi(s)$.

Theorem 2 below is an obvious, but useful, consequence of the string of inequalities in (2.6). The two results that follow it (Theorems 3 and 4) give simple characterizations of thrifty and equalizing plans, respectively.

Theorem 2 *A plan π is optimal at $s \in S$, if and only if π is both thrifty and equalizing at s.*

Theorem 3 *For a given plan π and initial state $s \in S$, the following are equivalent*:

(a) *the plan π is thrifty at s*;
(b) *the sequence $\{M_n, \mathcal{F}_n\}_{n\ge 1}$ is a martingale under $\mathbb{P}^{\pi,s}$; and*
(c) *for all $n \ge 1$, we have $\mathbb{P}^{\pi,s}(a_n$ conserves $V(\cdot)$ at $s_n) = 1$.*

Proof We write $\mathbb{E}[\,\cdot\,]$ for the expectation operator $\mathbb{E}^{\pi,s}[\,\cdot\,]$ below.

We start by assuming (a). Then, since $\mathbb{E}[M_n]$ is non-increasing in n, we have $\mathbb{E}[M_{n+1}] = \mathbb{E}[M_1] = V(s)$ for all $n \ge 1$. Hence, equality must hold in (2.5) with probability one, and (b) follows.

Now, let us assume (b). Then equality holds $\mathbb{P}^{\pi,s}$-almost surely in (2.5), and thus $(T_{a_n}V)(s_n) = V(s_n)$ almost surely, so (c) follows.

Finally, we assume (c). Taking expectations in (2.5), we see that $\mathbb{E}[M_{n+1}] = \mathbb{E}[M_n] = \mathbb{E}[M_1] = V(s)$ and, consequently $\Lambda^\pi(s) = V(s)$, so (a) follows. □

Theorem 4 *A given plan π is equalizing at $s \in S$, if and only if we have* $\lim_{n\to\infty}(\beta^n\,\mathbb{E}^{\pi,s}[V(s_{n+1})]) = 0$.

This result is an obvious consequence of (2.6). If the reward function admits an upper bound $0 \le r(s,a) \le K < \infty$ for all $s \in S$, $a \in A(s)$, then $0 \le V(s) \le K/(1-\beta)$ for all $s \in S$ and every plan π is equalizing.

Paraphrasing Blackwell [3], Theorem 3 says that a plan is thrifty if, with probability one, it makes no "immediate, irremediable mistakes" along any history; whereas Theorem 4 says that a plan is equalizing, if "it is certain to force the system into states where little further income can be anticipated."

We conclude this section with a brief look at the problem on a finite horizon. For $n = 1, 2, \dots$ and $s \in S$, define the *optimal n-day return* as

$$V_n(s) := \sup_{\pi} \mathbb{E}^{\pi,s}\left(\sum_{k=1}^{n} \beta^{k-1} r(s_k, a_k)\right). \tag{2.7}$$

The following result records the well-known backward induction algorithm and the fact that the optimal n-day return converges to that for the infinite-horizon problem. For a proof, see Sect. 9.5 of Bertsekas & Shreve [1].

Theorem 5 *Let $V_0(\cdot)$ be identically zero. Then for all $s \in S$ and $n = 1, 2, \dots$,*

(a) $V_{n+1}(s) = \sup_{a \in A(s)} (T_a V_n)(s)$, *and*
(b) $V(s) = \lim_{n\to\infty} V_n(s)$.

3 The Euler and Transversality Conditions

We now specialize to problems with concave daily reward functions and convex action sets as in Stokey & Lucas [12]. We shall use the notation and many of the assumptions of this book so, for brevity, will refer to it as just S&L.

As in S&L, we assume that the state space S is a product $S = X \times Z$, with a vector $s = (x, z)$ consisting of an "endogenous state" $x \in X$ and an "exogenous shock" $z \in Z$. The sets X and Z are nonempty Borel subsets of the Euclidean spaces $\mathbb{R}^l$ and $\mathbb{R}^k$, respectively; we shall assume that X is convex.

For each $s = (x, z)$ the *action set* $A(s) = \Gamma(x, z)$ is a nonempty Borel subset of X and is convex in x: that is, if $y \in \Gamma(x, z)$, $y' \in \Gamma(x', z)$, $z \in Z$, and $0 \le \theta \le 1$, then $\theta y + (1-\theta)y' \in \Gamma(\theta x + (1-\theta)x', z)$ (Assumption 9.11 of S&L).

The *daily reward function* is now of the form $r(s, y) = F(x, y, z)$; here $F : X \times X \times Z \to [0, \infty)$ is a given, Borel measurable "reward" function, concave in the pair (x, y), for every given $z \in Z$ (Assumption 9.10 of S&L).

The *law of motion* is of the form $s = (x, z) \longrightarrow (y, \mathfrak{z})$; here $a = y \in \Gamma(x, z)$ is the "action", and the distribution of the Z-valued random variable $\mathfrak{z}$ is given by a Markov kernel $\mathfrak{q}(d\xi|z)$. Note that the action y is the next value of the endogenous state: for $n \ge 1$ we have $y_n \equiv x_{n+1}$, a random variable measurable with respect to the σ-field

$$\mathcal{F}_n := \sigma(x_1, y_1, z_1, \dots, x_{n-1}, y_{n-1}, z_{n-1}, x_n, z_n). \tag{3.1}$$

In this setting, the Bellman equation becomes

$$V(x, z) = \sup_{y \in \Gamma(x,z)} \left(F(x, y, z) + \beta \int_Z V(y, \xi)\, \mathfrak{q}(d\xi|z)\right). \tag{3.2}$$

(There appears to be a technical oversight in S&L, pages 246 and 273, where it is stated that, under these conditions, there may not be a Bellman equation because of measurability issues; see Remark 1 above, and the Appendix, Theorem 14.) We shall denote the function occurring inside the supremum in (3.2) by

$$\psi(x,y,z) := F(x,y,z) + \beta \int_Z V(y,\xi)\,\mathfrak{q}(d\xi|z). \tag{3.3}$$

Lemma 6 *The value function $V(y,z)$ is concave in y; hence, so is $\psi(x,y,z)$. The function $\psi(x,y,z)$ is strictly concave in y, if $F(x,y,z)$ is.*

Proof Let $V_0(\cdot,\cdot)$ be identically zero and, for $n \geq 1$, let $V_n(\cdot,\cdot)$ be the optimal n-day return function as in (2.7) with $s=(x,z)$. Then, by Theorem 5(a),

$$V_{n+1}(x,z) = \sup_{y\in\Gamma(x,z)} \left(F(x,y,z) + \beta \int_Z V_n(y,\xi)\,\mathfrak{q}(d\xi|z) \right).$$

If, for each $z \in Z$, the function $V_n(\cdot,z)$ is concave, then $V_{n+1}(\cdot,z)$ is the supremum of a concave function over a convex set; by a well-known result, it is concave. An induction shows that all the $V_n(\cdot,z)$ are concave. From Theorem 5(b), the function $V(\cdot,z)$ is thus the pointwise limit of concave functions, and is therefore concave as well. The assertions about the function ψ of (3.3) are now easy to check. □

The usual treatment of the Euler and transversality conditions assumes that the plans in question are at interior states and choose interior actions. To be precise, we say that the state $s=(x,z)$ is *interior*, if x is in the interior of the set X; and we say that the action y at s is *interior*, if y belongs to the interior of the set $\Gamma(x,z)$. A plan π is called *interior* at $s=(x,z)$, if s is interior and, with probability one under $\mathbb{P}^{\pi,s}$, only interior states are visited and only interior actions are taken.

We assume for the rest of this section that the daily reward function $(x,y) \mapsto F(x,y,z)$ is continuous on $X \times X$ and continuously differentiable in the interior of $X \times X$, for every $z \in Z$. We shall use the notation $D_i F(x,y,z)$ for the partial derivative of F at (x,y,z) with respect to the ith coördinate, for $i=1,2,\ldots,2l$. Let $D_x F$ be the vector $(D_1 F, D_2 F, \ldots, D_l F)$ consisting of the partial derivatives of F with respect to its first l arguments, and let $D_y F$ be the vector $(D_{l+1} F, D_{l+2} F, \ldots, D_{2l} F)$ of the next l partial derivatives of F. We shall use similar notation for the partial derivatives of other functions, such as $V(x,z)$.

We shall assume (cf. Assumptions 9.8 and 9.9 in S&L) that the action sets $\Gamma(x,z)$ and the daily reward function $F(x,y,z)$ are nondecreasing in x; that is,

$$\Gamma(x,z) \subseteq \Gamma(x',z) \quad \text{and} \quad F(x,y,z) \leq F(x',y,z) \quad \text{whenever} \quad x \leq x'. \tag{3.4}$$

We shall also impose the following *continuity assumption* (cf. Assumption 9.6 in S&L): for every interior state (x_0,z) and interior action $y \in \Gamma(x_0,z)$, there exists an open neighborhood $\mathcal{O} \subset X$ of x_0 such that $y \in \Gamma(x,z)$, $\forall\, x \in \mathcal{O}$.

Lemma 7

(i) *The value function $V(x,z)$ is nondecreasing in x.*
(ii) *If the partial derivatives $D_x V(x,\xi)$ exist for $\mathfrak{q}(\cdot\,|z)$-almost all $\xi \in Z$, then, whenever both its sides are well-defined, the following equality holds*:

$$D_x \int_Z V(x,\xi)\,\mathfrak{q}(d\xi|z) = \int_Z D_x V(x,\xi)\,\mathfrak{q}(d\xi|z).$$

Proof To verify (i), let $x \le x'$ and consider any plan π for a player who begins at state (x,z). By (3.4), a second player at (x',z) can choose the same initial action y and receive an initial daily reward at least as large as that for the first player. Both players proceed to the state $(y,\mathfrak{z})$, where $\mathfrak{z}$ has distribution $\mathfrak{q}(\cdot|z)$. Thus the second player can earn the same rewards as the first thereafter.

For part (ii), consider for $\varepsilon > 0$ the quotients $(V(x_1,\dots,x_i+\varepsilon,\dots,x_l,z) - V(x_1,\dots, x_i,\dots,x_l,z))/\varepsilon$. By part (i), these are nonnegative; and by the concavity of $V(\cdot\,,z)$, they are nondecreasing as $\varepsilon \downarrow 0$ [10, pp. 4, 5]). The desired equality now follows by monotone convergence. □

Theorem 8 *Suppose that π is an interior plan at $s=(x,z)$. Then π is thrifty at s, if and only if the following hold with probability one under $\mathbb{P}^{\pi,s}$*:

(a) *the* Envelope equation

$$D_x V(x_n,z_n) = D_x F(x_n,y_n,z_n), \quad \forall n = 1,2,\dots;$$

(b) *the* Euler equation

$$D_y F(x_n,y_n,z_n) + \beta \int_Z D_x F(y_n,y_{n+1},\xi)\,\mathfrak{q}(d\xi|z_n) = 0, \quad \forall n = 1,2,\dots.$$

Proof Let us start by assuming that π is thrifty. By Theorem 3, the actions y_n conserve $V(\cdot,\cdot)$ at s_n for all $n \in \mathbb{N}$, on an event of probability one. Hence, for outcomes in this event, y_n maximizes $\psi(x_n,\cdot\,,z_n)$ over $\Gamma(x_n,z_n)$, and the envelope equation can be established exactly as in Theorem 9.10, page 267, of S&L; namely, using the concavity of $\psi(x_n,\,\cdot\,,z_n)$ from Lemma 6, and the fact that $D_x V(x_n,z_n) = D_x\psi(x_n,y_n,z_n)$ from Theorem 4.10, page 84 of S&L.

Here is the gist of this (pathwise) argument, reproduced here at the request of the referee: By the interiority and continuity assumptions there exists, for every $n \in \mathbb{N}$, an open neighborhood $\mathcal{O}_n$ of x_n such that $y_n \in \Gamma(x,z_n)$ holds for all $x \in \mathcal{O}_n$. Therefore, the function

$$W(x) := F(x,y_n,z_n) + \beta \int_Z V_n(y_n,\xi)\,\mathfrak{q}(d\xi|z_n) \tag{3.5}$$

is concave, continuously differentiable, and satisfies $W(x) \le V(x,z_n)$, $\forall\ x \in \mathcal{O}_n$, with equality for $x = x_n$. Now, any subgradient $p \in \mathbb{R}^l$ of the function $V(\cdot\,,z_n)$ at x_n must satisfy

$$p\cdot(x-x_n) \ge V(x,z_n) - V(x_n,z_n) \ge W(x) - W(x_n), \quad \forall x \in \mathcal{O}_n,$$

where the dot $\cdot$ signifies the usual inner product in $\mathbb{R}^l$. But $W(\cdot)$ is differentiable at x_n, so there is only one such subgradient; thus $V(\cdot, z_n)$ is differentiable at x_n (cf. [11, p. 242]) and we have $D_x V(x_n, z_n) = D_x W(x_n) = D_x F(x_n, y_n, z_n) = D_x \psi(x_n, y_n, z_n)$, as claimed.

The Euler equation, which can be written equivalently as

$$D_y F(x_n, y_n, z_n) + \beta \cdot \mathbb{E}^{\pi,s}\big[D_x F(y_n, y_{n+1}, z_{n+1}) \,|\, \mathcal{F}_{n+1}\big] = 0, \tag{3.6}$$

follows now by setting $D_y \psi(x_n, y, z_n) = 0$ at $y = y_n$, and recalling the envelope equation and part (ii) of Lemma 7. (Note that $(y, \xi) \mapsto D_x V(y, \xi)$ is continuous by (a).)

• For the converse, assume that (a) and (b) hold. We need to show that, with $\mathbb{P}^{\pi,s}$-probability one, y_n maximizes on the set $\Gamma(x_n, z_n)$ the concave function $\psi(x_n, \cdot, z_n)$ as in (3.3), for each $n \in \mathbb{N}$. But by (a), (b) and Lemma 7(ii), and recalling $y_n \equiv x_{n+1}$, we obtain, with $\mathbb{P}^{\pi,s}$-probability one:

$$\begin{aligned} D_y \psi(x_n, y_n, z_n) &= D_y F(x_n, y_n, z_n) + \beta \int_Z D_x V(y_n, \xi)\, \mathfrak{q}(d\xi | z_n) \\ &= D_y F(x_n, y_n, z_n) + \beta \int_Z D_x F(y_n, y_{n+1}, \xi)\, \mathfrak{q}(d\xi | z_n) = 0. \qquad \square \end{aligned}$$

To prove the necessity of the customary transversality condition for an optimal interior plan, we shall use both its thriftiness and equalization properties.

Theorem 9 *Suppose the plan π is optimal and interior at $s = (x, z)$, and that the reward function satisfies the requirement*

$$x \cdot D_x F(x, y, z) \ge 0 \tag{3.7}$$

for all interior states (x, z) and interior actions $y \in \Gamma(x, z)$. Then π satisfies

(c) *the* Transversality Condition

$$\lim_{n\to\infty} \Big(\beta^n \, \mathbb{E}^{\pi,s}\big[x_n \cdot D_x F(x_n, y_n, z_n) \big] \Big) = 0.$$

Proof Since π is optimal, it is thrifty by Theorem 2 and thus, from Theorem 8, satisfies with $\mathbb{P}^{\pi,s}$-probability one the envelope equation; therefore for all $n = 1, 2, \ldots$, we have the string of inequalities

$$\begin{aligned} V(x_n, z_n) &\ge V(x_n, z_n) - V(0, z_n) \\ &\ge x_n \cdot D_x V(x_n, z_n) = x_n \cdot D_x F(x_n, y_n, z_n) \ge 0. \end{aligned} \tag{3.8}$$

Here, the first inequality holds because $V(\cdot, \cdot)$ is nonnegative; the second inequality follows from a general property of concave functions (Theorem A, Chap. IV, page 98 in [10]); the equality is by the envelope equation (a) in Theorem 7; and the last inequality is from condition (3.7).

Since π is optimal, it is also equalizing by Theorem 2. Now take expectations under $\mathbb{P}^{\pi,s}$ in (3.8), and use Theorem 4. $\square$

Remark 2 Notice that $D_x F(x, y, z) \geq 0$ since, by (3.4), $F(x, y, z)$ is nondecreasing in x. Thus assumption (3.7) in the statement of Theorem 9 is automatically satisfied if all the states $x \in X$ lie in the nonnegative orthant $\mathbb{R}^l_+$ of $\mathbb{R}^l$, as is often true in economic applications. Furthermore, the assumption that the equality in Lemma 7(ii) holds was not needed in Theorem 8 to show that thriftiness implies the envelope equation. Consequently, this assumption is not needed in Theorem 9 to establish the transversality condition.

The next result is familiar to dynamic programmers working in economics.

Theorem 10 *Suppose the plan π is interior at $s = (x, z)$, and (3.7) holds.*

If π is optimal, then it satisfies both the Euler equation with $\mathbb{P}^{\pi,s}$-probability one, and the transversality condition. Conversely, if these two conditions hold for π and, in addition, we have $X \subseteq \mathbb{R}^l_+$, then π is optimal.

Proof If π is optimal, then it is thrifty by Theorem 2; if in addition it is interior at $s = (x, z)$, it satisfies also the Euler equations with $\mathbb{P}^{\pi,s}$-probability one, by Theorem 8. The transversality condition holds by Theorem 9.

• On the other hand, as stated in S&L, p. 281 it is straightforward to adapt their proof for the non-stochastic case (Theorem 4.15, p. 98), and show that the Euler equation and the Transversality condition together imply the optimality of π when $X \subseteq \mathbb{R}^l_+$. At the request of the referee, we sketch this adaptation.

Consider any other strategy $\widetilde{\pi}$ and suppose, as we may, that we have constructed on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ the state/action sequences $\{(x_n, y_n, z_n)\}_{n\in\mathbb{N}}$ and $\{(\widetilde{x}_n, \widetilde{y}_n, z_n)\}_{n\in\mathbb{N}}$ corresponding to π and $\widetilde{\pi}$, respectively, with $(\widetilde{x}_1, z_1) = (x_1, z_1) = s$; for notational convenience, let us also set $(x_0, y_0, z_0) \equiv (x_1, y_1, z_1)$. Then for the corresponding sequences $\{Q_n\}_{n\in\mathbb{N}}$ and $\{\widetilde{Q}_n\}_{n\in\mathbb{N}}$ as in (2.4),

$$\begin{aligned}
&\beta(Q_N - \widetilde{Q}_N)\\
&\quad= \sum_{n=1}^{N} \beta^n \big[F(x_n, y_n, z_n) - F(\widetilde{x}_n, \widetilde{y}_n, z_n)\big]\\
&\quad\geq \sum_{n=1}^{N} \beta^n \Big[(x_n - \widetilde{x}_n) \cdot D_x F(x_n, y_n, z_n) + (x_{n+1} - \widetilde{x}_{n+1}) \cdot D_y F(x_n, y_n, z_n)\Big]\\
&\quad= \sum_{n=1}^{N} \beta^{n-1} (x_n - \widetilde{x}_n) \cdot \Big[D_y F(x_{n-1}, y_{n-1}, z_{n-1}) + \beta\, D_x F(y_{n-1}, y_n, z_n)\Big]\\
&\qquad+ \beta^N (x_{N+1} - \widetilde{x}_{N+1}) \cdot D_y F(x_N, y_N, z_N), \quad \forall N \in \mathbb{N}
\end{aligned}$$

a.s. We have used $y_n \equiv x_{n+1}$, and the concavity of the function $F(\cdot, \cdot, z)$. Now we take expectations and recall the notation of (3.1), to obtain for all $N \in \mathbb{N}$:

$$\begin{aligned}
&\beta\left[\mathbb{E}(Q_N) - \mathbb{E}(\widetilde{Q}_N)\right]\\
&\quad\geq \beta^N\, \mathbb{E}\big[(x_{N+1} - \widetilde{x}_{N+1}) \cdot D_y F(x_N, y_N, z_N)\big]
\end{aligned}$$

$$+\,\mathbb{E}\sum_{n=0}^{N-1}\beta^n(x_{n+1}-\widetilde{x}_{n+1})$$
$$\times\Big[\,D_yF(x_n,y_n,z_n)+\beta\,\mathbb{E}\big(D_xF(y_n,y_{n+1},z_{n+1})\,|\,\mathcal{F}_{n+1}\big)\Big].$$

Thanks to the version (3.6) of the Euler equation, this last (summation) term is equal to zero; as for the term that precedes it, this same equation allows us to write it as

$$\begin{aligned}&\beta^{N+1}\,\mathbb{E}\big[\,(\widetilde{x}_{N+1}-x_{N+1})\cdot D_xF(y_N,y_{N+1},z_{N+1})\,\big]\\&\quad\ge-\beta^{N+1}\,\mathbb{E}\big[\,x_{N+1}\cdot D_xF(x_{N+1},y_{N+1},z_{N+1})\,\big].\end{aligned}$$

For this last inequality, we have used the assumption $X\subseteq\mathbb{R}^l_+$ and recalled Remark 2. We conclude from all this:

$$\beta^N\,\mathbb{E}\big[\,x_{N+1}\cdot D_xF(x_{N+1},y_{N+1},z_{N+1})\,\big]+\mathbb{E}(Q_N)\ge\mathbb{E}(\widetilde{Q}_N),\quad\forall N\in\mathbb{N}.$$

Passing to the limit as $N\to\infty$ we obtain, with the help of the transversality condition and monotone convergence, the comparison $R^\pi(s)\ge R^{\widetilde{\pi}}(s)$ in the notation of (2.2), that is, the optimality of π. □

By Theorems 2 and 10, the thriftiness and equalization conditions are equivalent to the Euler equations and the transversality condition for the special class of problems of this section, when the plan π is interior.

4 Envelope and Euler Inequalities

In the one-dimensional case $l=1$ it is possible to replace the Envelope Equation and the Euler Equation of Theorem 8 with appropriate inequalities, thereby dispensing with interiority assumptions on the part of the plan π.

Let us illustrate this possibility by taking

$$X=[0,\infty),\qquad\Gamma(x,z)=\big[0,\gamma(x,z)\big),$$

where $\gamma:[0,\infty)\times Z\to[0,\infty)$ is continuous, concave, and non-decreasing in the first argument: $\Gamma(x,z)\subseteq\Gamma(x+\varepsilon,z)$ holds for every $\varepsilon>0$ and $(x,z)\in[0,\infty)\times Z$. We also assume that $F(\cdot\,,y,z)$ is concave, nondecreasing, and nonnegative for all (y,z), but need no longer assume that this function is differentiable.

We shall denote by $D_x^\pm F(x_0,y,z)$, $D_x^\pm V(x_0,z)$ the left- and right-derivatives at $x=x_0$ of the concave functions $F(\cdot\,,y,z)$ and $V(\cdot\,,z)$, respectively.

Theorem 11 (Envelope Inequalities) *If the plan π is thrifty at $s=(x_1,z_1)$ then, with probability one under $\mathbb{P}^{\pi,s}$, we have for all $n=1,2,\ldots$ the properties*

$$D_x^+V(x_n,z_n)\ge D_x^+F(x_n,y_n,z_n),\tag{4.1}$$

$$D_x^-V(x_n,z_n)\le D_x^-F(x_n,y_n,z_n)\quad\textit{on}\ \{y_n<\gamma(x_n,z_n)\}.\tag{4.2}$$

Proof By Theorem 3, the actions y_n conserve $V(\cdot,\cdot)$ at (x_n, z_n) on an event of $\mathbb{P}^{\pi,s}$-probability one, so we have $W(x_n) = V(x_n, z_n)$ for the function $W(\cdot)$ of (3.5). Furthermore, $y_n \in \Gamma(x_n, z_n) \subseteq \Gamma(x_n + \varepsilon, z_n)$ holds for $\varepsilon > 0$, hence $W(x_n + \varepsilon) \leq V(x_n + \varepsilon, z_n)$, and (4.1) follows.

As for the second inequality, it follows from the continuity of the function $\gamma(\cdot\,,\, z_n)$ that $y_n \in \Gamma(x_n - \varepsilon, z_n)$ for $\varepsilon > 0$ sufficiently small. Hence, for such ε, $W(x_n - \varepsilon) \leq V(x_n - \varepsilon, z_n)$ and (4.2) follows. □

Theorem 12 (Euler Inequalities) *If the plan π is thrifty at $s = (x_1, z_1)$ then, with probability one under $\mathbb{P}^{\pi,s}$, we have for all $n = 1, 2, \ldots$ the inequalities*

$$0 \geq D_y^+ F(x_n, y_n, z_n) + \beta \int_Z D_x^+ F(y_n, y_{n+1}, \xi)\, \mathfrak{q}(d\xi|z_n) \quad \textit{on } \{y_n < \gamma(x_n, z_n)\}, \tag{4.3}$$

$$0 \leq D_y^- F(x_n, y_n, z_n) + \beta \int_Z D_x^- F(y_n, y_{n+1}, \xi)\, \mathfrak{q}(d\xi|z_n) \quad \textit{on } \{y_n > 0\}. \tag{4.4}$$

Proof By Theorem 3, with probability one under $\mathbb{P}^{\pi,s}$, the action y_n conserves $V(\cdot,\cdot)$ at each state $s_n = (x_n, z_n)$, $n = 1, 2, \ldots$. To wit, the concave function $y \mapsto \psi(x_n, y, z_n)$ of (3.3) is maximized over $\Gamma(x_n, z_n)$ at y_n. This implies

$$0 \geq D_y^+ \psi(x_n, y_n, z_n) \quad \text{provided} \quad y_n < \gamma(x_n, z_n) \tag{4.5}$$

as well as

$$0 \leq D_y^- \psi(x_n, y_n, z_n) \quad \text{provided} \quad y_n > 0. \tag{4.6}$$

For each $(y, z) \in [0, \infty) \times Z$, the quotients $(V(y + \varepsilon, z) - V(y, z))/\varepsilon$, $\varepsilon \geq 0$ are nonnegative, and increase as $\varepsilon \downarrow 0$. This is because the function $F(\cdot, y, z)$ is increasing and concave, which implies that $V(\cdot, z)$ is also increasing and concave. Thus, by monotone convergence, we obtain

$$D_y^+ \int_Z V(y, \xi)\, \mathfrak{q}(d\xi|z) = \int_Z D_y^+ V(y, z)\, \mathfrak{q}(d\xi|z).$$

Similar reasoning gives the same formula for left-derivatives at $(y, z) \in (0, \infty)$.

An application of the Envelope Inequality (4.1) to (4.5) now yields

$$\begin{aligned} 0 \geq D_y^+ \psi(x_n, y_n, z_n) &= D_y^+ F(x_n, y_n, z_n) + \beta \int_Z D_y^+ V(y_n, \xi)\, \mathfrak{q}(d\xi|z_n) \\ &\geq D_y^+ F(x_n, y_n, z_n) + \beta \int_Z D_x^+ F(y_n, y_{n+1}, \xi)\, \mathfrak{q}(d\xi|z_n) \end{aligned}$$

on $\{y_n < \gamma(x_n, z_n)\}$. This establishes (4.3); and (4.4) is proved similarly. □

It is now possible, in the special setting of this section, to show that a transversality condition is necessary for a plan to be optimal, even without interiority or smoothness of the daily reward.

Theorem 13 (Transversality Condition) *If π is optimal at $s=(x_1,z_1)$, then*

$$\lim_{n\to\infty}\left(\beta^n\,\mathbb{E}^{\pi,s}[x_n\cdot D_x^+F(x_n,y_n,z_n)]\right)=0. \tag{4.7}$$

Proof We calculate as follows:

$$\begin{aligned}V(x_n,z_n)&\ge V(x_n,z_n)-V(0,z_n)=\int_0^{x_n}D_x^+V(x,z_n)\,dx\\&\ge x_n\cdot D_x^+V(x_n,z_n)\ge x_n\cdot D_x^+F(x_n,y_n,z_n)\ge 0.\end{aligned}$$

Here the equality follows from a general property of concave functions (Problem A, p. 13 in [10]). The second inequality holds because $D_x^+V(x,z_n)$ is nonincreasing in x; the third is by the Envelope inequality (4.1), which applies because the optimality of π implies its thriftiness by Theorem 2; the final inequality holds because $x_n\ge 0$ and $F(\cdot,y_n,z_n)$ is nondecreasing. Since π is optimal, it is equalizing by Theorem 2. Now apply Theorem 4. □

We leave open the question of whether there is a converse in the context of this section. That is, if a plan π satisfies the Transversality Condition (4.7) and the Euler Inequalities (4.3), (4.4) hold with probability one, is π then optimal?

Appendix: Measurability

Our objective here is to describe a fairly general class of dynamic programming problems for which the optimal reward function is measurable in an appropriate sense. We shall only sketch the proof and provide references for further details.

A dynamic programming problem (S,A,q,r,β) as in Sect. 2 will be called *measurable*, if the following hold:

(a) The state space S is a nonempty Borel subset of a Polish space. (A topological space is called *Polish*, if it is homeomorphic to a complete, separable metric space. In particular, any Euclidean space is Polish.)
(b) There is a Polish space X that contains the union of the action sets $A(s)$, $s\in S$. Furthermore, the set $\widetilde{A}$ below is a Borel subset of the product space $S\times X$:

$$\widetilde{A}:=\{(s,a):s\in S,\ a\in A(s)\}.$$

(c) The law of motion q is a Borel-measurable *transition function* from $\widetilde{A}$ to S. That is, for each fixed $(s,a)\in\widetilde{A}$, $q(\cdot\,|s,a)$ is a probability measure on the Borel subsets of S; and for each fixed Borel subset B of S, $q(B\,|\cdot,\cdot)$ is a Borel-measurable function on $\widetilde{A}$.
(d) The daily reward function $r:S\times X\to[0,\infty)$ is Borel measurable.

We also need to impose measurability conditions on the plans that a player is allowed to choose. To do so, we introduce the notion of *universal measurability*.

Let Y be a Polish space and let $\mathcal{B}$ be its σ-field of Borel subsets.

Definition 3 A subset U of Y is called *universally measurable,* if it belongs to the completion of $\mathcal{B}$ under every probability measure μ on $\mathcal{B}$.

The set of all universally measurable subsets of Y is a σ-field $\mathcal{U}$ larger than $\mathcal{B}$, if Y is uncountable. A function $f : Y \to Z$, where Z is another Polish space, is called *universally measurable,* if $f^{-1}(C) \in \mathcal{U}$ holds for every Borel subset C of Z. Notice that $\int f\, d\mu$ is well-defined for every universally measurable function $f : Y \to [0, \infty)$ and every probability measure μ defined on $\mathcal{B}$.

A plan $\pi = (\pi_1, \pi_2, \ldots)$ is *universally measurable* if, for every $n = 1, 2, \ldots, \pi_n$ is a universally measurable function from $(S \times X)^{n-1} \times S$ to X. Let Π be the set of all universally measurable plans π. Let $H = S \times X \times S \times X \times \cdots$ be the Polish space of all infinite histories $h = (s_1, a_1, s_2, a_2, \ldots)$. Each state $s \in S$ together with a plan $\pi \in \Pi$ determines a probability measure $\mathbb{P}^{\pi,s}$ on the Borel subsets of H. The optimal reward $V(s)$ at $s \in S$ is now defined by

$$V(s) := \sup_{\pi \in \Pi} R^{\pi}(s) = \sup_{\pi \in \Pi} \int \mathfrak{g}(h)\, d\,\mathbb{P}^{\pi,s}(h);$$

here $\mathfrak{g}(\cdot)$ is the Borel-measurable function defined for $h \in H$ by

$$\mathfrak{g}(h) := \mathfrak{g}(s_1, a_1, s_2, a_2, \ldots) = \sum_{n=1}^{\infty} \beta^{n-1} r(s_n, a_n). \tag{A.1}$$

Theorem 14 ([13]) *The optimal reward function $V(\cdot)$ of a measurable dynamic programming problem is universally measurable.*

Proof We will only sketch the main ideas; for more details we refer the reader to Theorem 4.2 of Feinberg [5]. (It should be noted that Feinberg uses Borel rather than universally measurable plans and, for this reason, must assume that the set $\widetilde{A}$ contains the graph of a Borel-measurable function from S into X.)

Let $\mathcal{M}(H)$ be the set of all probability measures on the Borel subsets of H. Then $\mathcal{M}(H)$, when equipped with its usual topology of vague convergence, is again a Polish space (cf. [8]). It can be shown that

$$\mathcal{L} := \left\{(s, \mathbb{P}^{\pi,s}) : s \in S,\ \pi \in \Pi\right\}$$

is a Borel subset of $S \times \mathcal{M}(H)$ (see Sect. 3 of [5]). For $s \in S$, let $\mathcal{L}(s)$ be the s-section of $\mathcal{L}$; that is,

$$\mathcal{L}(s) = \left\{\mu \in \mathcal{M}(H) : \mu = \mathbb{P}^{\pi,s} \text{ for some } \pi \in \Pi\right\}.$$

Then, with $\mathfrak{g}(\cdot)$ as in (A.1), we have $V(s) = \sup\{\int \mathfrak{g} d\mu : \mu \in \mathcal{L}(s)\},\ s \in S$. It is not difficult to check that the function $\mathcal{M}(H) \ni \mu \mapsto \int \mathfrak{g}\, d\mu \in \mathbb{R}$ is Borel measurable. Also, for each $c \in \mathbb{R}$, the set $S_c = \{s \in S : V(s) > c\}$ is the projection onto S of the Borel set $B_c = \{(s, \mu) \in \mathcal{L} : \int \mathfrak{g}\, d\mu > c\}$. Thus S_c is an *analytic* set, and therefore universally measurable (Corollary 7.42.1, p. 169 in [1]). It follows that $V(\cdot)$ is universally measurable. $\square$

The assumption made in this paper that the daily reward function $r(\cdot, \cdot)$ is nonnegative is not necessary for the proof of Theorem 14 or for the proof of the Bellman equation. See [1, 13], and [5] for more general results.

References

1. Bertsekas, D.P., Shreve, S.E.: Stochastic Optimal Control: The Discrete-Time Case. Academic Press, New York (1978)
2. Blackwell, D.: Discounted dynamic programming. Ann. Math. Stat. **36**, 226–235 (1965)
3. Blackwell, D.: On stationary policies. J. R. Stat. Soc. A **133**, 33–37 (1970)
4. Dubins, L.E., Savage, L.J.: How to Gamble if You Must: Inequalities for Stochastic Processes. McGraw-Hill, New York (1965)
5. Feinberg, E.A.: On measurability and representation of strategic measures in Markov decision processes. In: Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell. Institute of Mathematical Statistics, Lecture Notes—Monograph Series, vol. 30, pp. 29–43 (1996)
6. Hordijk, A.: Dynamic Programming and Markov Potential Theory. Mathematical Centre Tracts 51, Amsterdam (1974)
7. Kamihigashi, T.: Necessity of the transversality condition for stochastic models with bounded or CRRA utility. J. Econ. Dyn. Control **29**, 1313–1329 (2005)
8. Parthasarathy, K.R.: Probability Measures on Metric Spaces. Academic Press, New York (1967)
9. Rieder, U.: On optimal policies and martingales in dynamic programming. J. Appl. Probab. **13**, 507–518 (1976)
10. Roberts, A.W., Varberg, D.E.: Convex Functions. Academic Press, New York (1973)
11. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
12. Stokey, N.L., Lucas, R.E.: Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge (1989)
13. Strauch, R.E.: Negative dynamic programming. Ann. Math. Stat. **37**, 871–890 (1966)