Two Characterizations of Optimality in Dynamic Programming *

Ioannis Karatzas and Wiliam D. Sudderth

May 9, 2008

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.

1 Introduction

It was shown by Dubins and Savage (1965) 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 Blackwell (1970), Hordijk (1974), Reider (1976) and Blume et al. (1982), 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 and Lucas (1989) 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 and Lucas, page 102, and Kamihigashi (2005).)

The notions of "thrifty" and "equalizing" seem not to be widely known to dynamic programmers working in economics, although they have proved to be quite useful in other contexts. We hope that this note 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

*Research supported in part by the National Science Foundation, under Grant NSF-DMS-06-01774. The first author wishes to thank the Mathematics and Statistics Departments of the University of Minnesota, for their hospitality during a brief visit in May 2008.

to the thrifty and equalizing conditions. In Section 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 theory 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 $s \in S,\ a \in A(s)$ 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

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

of states and actions, along with 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 n^{th} action a_n as a function $\pi_n(h_n)$ of the previous history $h_n = (s_1, a_1, \cdots, s_{n-1}, a_{n-1}, s_n)$. 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, which we write as

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

The *optimal reward* or *value* at 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 (Blackwell, 1965). 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. $\square$

We shall assume, for simplicity, that

$$V(s) < \infty\,, \quad \forall \;\; s \in S\,.$$

A key tool is the *Bellman equation*

$$V(s) \;=\; \sup_{a \in A(s)} \left(r(s,a) + \beta \int_S V(\sigma)\, q(d\sigma|s,a) \right),$$

which holds in great generality and is also known as the "optimality equation" (see, for example, section 9.4 of Bertsekas and Shreve, 1978). For $a \in A(s)$ and a measurable function $x : S \mapsto \mathbb{R}^+$, define

$$(T_a x)(s) \;:=\; r(s,a) + \beta \int_S x(\sigma)\, q(d\sigma|s,a).$$

The Bellman equation can now be written in the form

$$V(s) = \sup_{a \in A(s)} \big[\, (T_a V)(s) \,\big]\,, \qquad s \in S\,.$$

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

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

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 $\{Q_n\}_{n \ge 1}$, where

$$Q_n := \sum_{k=1}^{n} \beta^{k-1} r(s_k, a_k)\,, \tag{2.2}$$

and

$$M_1 := V(s_1), \qquad M_{n+1} := Q_n + \beta^n V(s_{n+1}), \;\; n \ge 1. \tag{2.3}$$

Let $\mathcal{F}_n$ be the $\sigma-$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 $n \ge 1$ and 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

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

$$\mathbb{E}^{\pi,s}[M_{n+1}|\mathcal{F}_n] = Q_{n-1} + \beta^{n-1} \cdot (T_{a_n} V)(s_n) \le Q_{n-1} + \beta^{n-1} \cdot V(s_n) = M_n \tag{2.4}$$

almost surely under $\mathbb{P}^{\pi,s}$. 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 (2.3), 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. $\square$

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

Define

$$\Lambda^\pi(s) := \lim_n \mathbb{E}^{\pi,s}(M_n).$$

Then

$$\begin{aligned} V(s) &= \mathbb{E}^{\pi,s}(M_1) \geq \lim_n \mathbb{E}^{\pi,s}(M_{n+1}) = \Lambda^\pi(s) \\ &= \lim_n \left\{\mathbb{E}^{\pi,s}(Q_n) + \beta^n\, \mathbb{E}^{\pi,s}[V(s_{n+1})]\right\} \\ &= R^\pi(s) + \lim_n \left\{\beta^n\, \mathbb{E}^{\pi,s}[V(s_{n+1})]\right\} \\ &\geq R^\pi(s)\,. \end{aligned} \tag{2.5}$$

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

Here is an obvious, but useful, consequence of the string of inequalities in (2.5).

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

The next two results give simple characterizations of thrifty and equalizing plans, respectively.

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 $\{M_n, \mathcal{F}_n\}_{n\geq 1}$ is a martingale under $\mathbb{P}^{\pi,s}$; and

(c) for all $n \geq 1$, we have $\mathbb{P}^{\pi,s}\big(a_n$ conserves $V(\cdot)$ at $s_n\big) = 1$.

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

Assume *(a)*. Then, since $\mathbb{E}[M_n]$ is non-increasing in n, we have $\mathbb{E}[M_{n+1}] = \mathbb{E}[M_n] = \mathbb{E}[M_1] = V(s)$ for all $n \geq 1$. Hence, equality must hold in (2.4) with probability one, and *(b)* follows.

Now assume *(b)*. Then equality holds $\mathbb{P}^{\pi,s}-$almost surely in (2.4), and thus $(T_{a_n}V)(s_n) = V(s_n)$ almost surely, so *(c)* follows.

Finally, assume *(c)*. Taking expectations in (2.4), we see that $\mathbb{E}[M_{n+1}] = \mathbb{E}[M_n] = \cdots = \mathbb{E}[M_1] = V(s)$ and, consequently $\Lambda^\pi(s) = V(s)$, so *(a)* follows. □

The next result is obvious from (2.5).

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

Paraphrasing Blackwell (1970), 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

$$V_n(s) \;:=\; \sup_{\pi} \mathbb{E}^{\pi,s}\left(\sum_{k=1}^{n} \beta^{k-1} r(s_k, a_k)\right). \tag{2.6}$$

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 section 9.5 of Bertsekas and Shreve (1978).

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) $V(s) = \lim_n V_n(s)$.

3 The Euler and transversality conditions

We now specialize to problems with concave daily reward functions and convex action sets as in Stokey and Lucas (1989). We shall use the notation and many of the assumptions of Stokey and Lucas and, for the sake of brevity, will refer to their book as just S&L.

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

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 $y \in \Gamma(x, z)$, $y' \in \Gamma(x', z)$, and $0 \le \theta \le 1$, then we have $\theta y + (1-\theta) y' \in \Gamma(\theta x + (1-\theta) x', z)$, as in 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

$$s = (x, z) \longrightarrow (y, \mathfrak{z}),$$

where $a = y \in \Gamma(x, z)$ 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: $y_n \equiv x_{n+1}$ for $n \ge 1$. The Bellman equation becomes

$$V(x, z) \;=\; \sup_{y \in \Gamma(x,z)} \left(F(x, y, z) + \beta \int_Z V(y, \xi)\, \mathfrak{q}(d\xi|z)\right). \tag{3.1}$$

(There is 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, as well as the Appendix.) Let

$$\psi(x, y, z) := F(x, y, z) + \beta \int_Z V(y, \xi)\, \mathfrak{q}(d\xi|z) \tag{3.2}$$

be the function occuring inside the supremum in (3.1).

Lemma 6. *The value function* $V(y, z)$ *is concave in* y*, and, 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 (2.6) 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 $V_n(\cdot\,, z)$ is concave, then $V_{n+1}(\cdot\,, z)$ is the supremum of a concave function over a convex set and, by a well-known result, is also concave. By induction, we conclude that all of the functions $V_n(\cdot\,, z)$ are concave. By Theorem 5(b), the function $V(\cdot\,, z)$ is the pointwise limit of concave functions and is therefore concave as well. The assertions about the function ψ of (3.2) 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 remainder of this section that the daily reward function $F(x, y, z)$ is continuous on $X \times X \times Z$, and continuously differentiable in x and y for (x, y) in the interior of $X \times X$. 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 i^{th} 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,

$$\Gamma(x, z) \subseteq \Gamma(x', z) \quad \text{and} \quad F(x, y, z) \leq F(x', y, z) \quad \text{whenever} \quad x \leq x'. \tag{3.3}$$

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*

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

Proof. To verify *(i),* let $x \le x'$ and consider any plan π for a player who begins at state (x,z). By (3.3) 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 the quotients

$$\frac{1}{\varepsilon}\cdot\Big(V(x_1,\ldots,x_i+\varepsilon,\ldots,x_l,z) - V(x_1,\ldots,x_i,\ldots,x_l,z)\Big)\,, \qquad \text{for } \ \varepsilon > 0\,.$$

By part *(i),* these quotients are nonnegative; and by the concavity of $V(\cdot\,,z)$, they are nondecreasing as $\varepsilon \downarrow 0$ (see Roberts and Varberg (1973), pages 4 and 5). The desired equality now follows from the monotone convergence theorem. $\square$

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* $\mathbb{P}^{\pi,s}$*, for all* $n = 1, 2, \ldots$*:*

(a) the Envelope equation:

$$D_x V(x_n,z_n) \;=\; D_x F(x_n,y_n,z_n),$$

(b) the Euler equation:

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

Proof. By Theorem 3 the actions y_n conserve $V(\cdot,\cdot)$ at s_n with probability one. Hence y_n maximizes $\psi(x_n,\cdot\,,z_n)$ over $\Gamma(x_n,z_n)$ on an event of probability one. The envelope equation can now be proved for outcomes in this event exactly as in the proof of 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. The Euler equation follows 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).*)

To prove the converse, assume that *(a)* and *(b)* hold. We need to show that, with $\mathbb{P}^{\pi,s}-$probability one, y_n maximizes the concave function

$$y \;\longmapsto\; \psi(x_n,y,z_n) \;=\; F(x_n,y,z_n) + \beta\int_Z V(y,\xi)\,\mathfrak{q}(d\xi|z_n)$$

on the set $\Gamma(x_n,z_n)$, for each $n\in\mathbb{N}$. But by *(a)*, *(b)* and Lemma 7(ii), we obtain

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

with $\mathbb{P}^{\pi,s}-$probability one. $\square$

To prove the necessity of the customary transversality condition for an optimal interior plan, we shall make use of 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*

$$x \cdot D_x F(x, y, z) \geq 0 \tag{3.4}$$

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

$$\lim_n \left(\beta^n \, \mathbb{E}^{\pi, s} \left[x_n \cdot D_x F(x_n, y_n, z_n) \right] \right) = 0 \,. \tag{3.5}$$

(The dot $\cdot$ in (3.4) and (3.5) signifies the usual inner product in $\mathbb{R}^l$.)

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

$$\begin{aligned} V(x_n, z_n) &\geq V(x_n, z_n) - V(0, z_n) \\ &\geq x_n \cdot D_x V(x_n, z_n) = x_n \cdot D_x F(x_n, y_n, z_n) \geq 0 \,. \end{aligned} \tag{3.6}$$

Here, the first inequality holds because $V(\cdot\,,\cdot)$ is nonnegative; the second inequality follows from a general property of concave functions (Theorem A, Chapter IV, page 98 in Roberts and Varberg (1973)); the equality is by the envelope equation (a) in Theorem 7; and the last inequality is from condition (3.4).

Since π is optimal, it is also equalizing by Theorem 2. Now take expectations under $\mathbb{P}^{\pi,s}$ in (3.6), and use Theorem 4. □

Notice that $D_x F(x, y, z) \geq 0$ since, by (3.3), $F(x, y, z)$ is nondecreasing in x. Thus assumption (3.4) in the statement of Theorem 9 is automatically true if all the states $x \in X$ lie in the nonnegative orthant of $\mathbb{R}^l$, as is often true in economic applications.

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

Theorem 10. *Suppose the plan π is interior at $s = (x, z)$, and that the assumption (3.4) holds. Then π is optimal, if and only if it satisfies both the Euler equation with $\mathbb{P}^{\pi,s}-$probability one, and the transversality condition.*

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.

As stated in S&L, page 281, it is straightforward to adapt their proof for the non-stochastic case (Theorem 4.15, page 98) that the two conditions of Euler and transversality imply optimality. □

By Theorems 2 and 9, the thrifty and equalizing conditions are equivalent to the Euler equations and the transversality condition for the special 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. In particular, we have $\Gamma(x, z) \subseteq \Gamma(x+\varepsilon, z)$ for $\varepsilon > 0$ and every $(x, z) \in [0, \infty) \times Z$. We also assume that $F(\cdot, y, z)$ is concave, nondecreasing, and nonnegative for all (x, z), but we shall no longer assume that this function is differentiable.

We shall denote by $D_x^{\pm} F(x_0, y, z)$ and $D_x^{\pm} V(x_0, z)$ 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 $P^{\pi,s}$, we have for all $n = 1, 2, \ldots$ the properties*

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

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

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. Consider the function

$$x \longmapsto W(x) := F(x, y_n, z_n) + \beta \int_Z V(y_n, \xi)\, \mathfrak{q}(d\xi|z_n).$$

Now $W(x_n) = V(x_n, z_n)$ when y_n conserves $V(\cdot, \cdot)$ at (x_n, z_n). Furthermore, $y_n \in \Gamma(x_n, z_n) \subseteq \Gamma(x_n+\varepsilon, z_n)$ holds for $\varepsilon > 0$, hence $W(x_n+\varepsilon) \leq V(x_n+\varepsilon, z_n)$. The inequality (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, $W(x_n - \varepsilon) \leq V(x_n - \varepsilon, z_n)$. The inequality (4.2) follows. $\square$

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*

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

and

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

Proof. By Theorem 3, we have with probability one that 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 (3.2) is maximized over $\Gamma(x_n, z_n)$ at y_n. This implies

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

as well as

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

For each $(y, z) \in [0, \infty) \times Z$, the quotients

$$\frac{1}{\varepsilon} \cdot \Big(V(y + \varepsilon, z) - V(y, z) \Big), \qquad \varepsilon > 0$$

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 the event $\{y_n < \gamma(x_n, z_n)\}$. This proves (4.3); the inequality (4.4) is proved similarly. $\square$

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*

$$\lim_n \Big(\beta^n\, \mathbb{E}^{\pi,s} [\, x_n \cdot D_x^+ F(x_n, y_n, z_n)\,] \Big) \;=\; 0\,. \tag{4.7}$$

Proof. We calculate as follows:

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

The equality above follows from a general fact about concave functions (see problem A, page 13 in Roberts and Varberg (1973)); the second inequality holds because $D_x^+ V(x, z_n)$ is nonincreasing in x; the third inequality is by the Envelope inequality (4.1), which applies because the optimality of π implies it is thrifty by Theorem 2; the final inequality holds because $x_n \geq 0$ and $F(\cdot, y_n, z_n)$ is nondecreasing. Since π is optimal, it is equalizing by Theorem 2. Now apply Theorem 4. $\square$

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) and (4.4) hold with probability one, is π then necessarily optimal?

5 Appendix on Measurability

Our object 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 section 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 *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

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

is a Borel subset of the product space $S \times X$.

(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 is a Borel-measurable function from $S \times X$ to $[0, \infty)$.

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 collection of all universally measurable subsets of Y forms a σ-field $\mathcal{U}$ which is 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 $n = 1, 2, \ldots$, π_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

$$\mathfrak{g}(h) := \mathfrak{g}(s_1, a_1, s_2, a_2, \ldots) = \sum_{n=1}^{\infty} \beta^{n-1} r(s_n, a_n)\,. \tag{5.1}$$

Theorem 14. (Strauch, 1966) *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, see Theorem 4.2 of Feinberg (1996). (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. Parthasarathy, 1967). It can be shown that the set

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

is a Borel subset of $S \times \mathcal{M}(H)$ (see section 3 of Feinberg (1996)). 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 (5.1), we have

$$V(s) = \sup \left\{ \int \mathfrak{g}\, d\mu \,:\, \mu \in \mathcal{L}(s) \right\}, \quad s \in S\,.$$

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 real number c, 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, page 169 in Bertsekas and Shreve (1978); Theorem 10.40, page 393 in Aliprantis and Border (1999)). It follows that $V(\cdot)$ is universally measurable. □

We imposed throughout this paper the assumption that the daily reward function $r(\cdot\,,\cdot)$ is nonnegative. This assumption is not necessary for the proof of Theorem 14, or for the proof of the Bellman equation; see Strauch (1966), Bertsekas and Shreve (1978), and Feinberg (1996) for more general results.

References

Aliprantis, C. D. & K. C. Border (1999). *Infinite-Dimensional Analysis: A Hitchhiker's Guide.* Second Edition, Springer-Verlag, New York.

Blume, L., D. Easley & M. O'Hara (1982) Characterization of optimal plans for stochastic dynamic programs. *Journal of Economic Theory* **28**, 2212–234.

Dubins, L. E. & L. J. Savage (1965). *How to Gamble if You Must: Inequalities for Stochastic Processes.* McGraw-Hill, New York.

Bertsekas, D. P. & S. E. Shreve (1978). *Stochastic Optimal Control: The Discrete-Time Case.* Academic Press, New York.

Blackwell, D. (1965). Discounted dynamic programming. *Annals of Mathematical Statistics* **36**, 226-235.

Blackwell, D. (1970). On stationary policies. *Journal of the Royal Statistical Society* **133A**, 33-37.

Feinberg, E. A. (1996). On measurability and representation of strategic measures in Markov decision processes. *Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell.* Institute of Mathematical Statistics: Lecture Notes - Monograph Series **30**, 29-43.

Hordijk, A. (1974). *Dynamic Programming and Markov Potential Theory.* Mathematical Centre Tracts **51**, Amsterdam.

Kamihigashi, T. (2005). Necessity of the transversality condition for stochastic models with bounded or CRRA utility. *Journal of Economic Dynamics & Control* **29**, 1313-1329.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

Rieder, U. (1976). On optimal policies and martingales in dynamic programming. *Journal of Applied Probability* **13**, 507-518.

Roberts, A. W. & D. E. Varberg (1973). *Convex Functions.* Academic Press, New York.

Stokey, N. L. & R. E. Lucas (1989). *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, MA.

Strauch, R. E. (1966). Negative dynamic programming. *Annals of Mathematical Statistics* **37**, 871-890.

Ioannis Karatzas
Department of Mathematics
Columbia University
New York, New York 10027
ik@math.columbia.edu

William D. Sudderth
School of Statistics
University of Minnesota
Minneapolis, Minnesota 55455
bill@stat.umn.edu