FROM THE BOOK BY D. BERTSEKAS
Dynamic Programming and Optimal Control, Vol. II, (4th Edition), Athena Scientific, 2012.

APPENDIX A:

Measure Theoretic Issues in Dynamic Programming

A general theory of stochastic dynamic programming must deal with the formidable mathematical questions that arise from the presence of uncountable probability spaces. The purpose of this appendix is to orient the mathematically advanced reader on these questions.†

The appendix is based on the research monograph by Bertsekas and Shreve [BeS78] (freely available from the internet), to which we refer for a detailed analysis, for references to earlier research, and for the development of mathematical background and terminology on Borel spaces and related subjects. We will explore here the main questions by means of a simple two-stage example described in Section A.1. In Section A.2, we develop a framework, based on universally measurable policies, for the rigorous mathematical development of the standard DP results for this example and for more general finite horizon models.

A.1 A TWO-STAGE EXAMPLE

Suppose that the initial state x_0 is a point on the real line $\Re$. Knowing x_0, we must choose a control $u_0 \in \Re$. Then the new state x_1 is generated

† The style and terminology of this appendix assume a reader who has knowledge of the basic notions of measure theory and is also familiar with finite horizon DP. In particular, we freely use basic notions of measurability and integration. We also use "inf" notation rather than "min" in various optimization equations, when the infimum is not known to be attained.

according to a transition probability measure $p(dx_1 \mid x_0, u_0)$ on the Borel σ-algebra of $\Re$ (the one generated by the open sets of $\Re$). Then, knowing x_1, we must choose a control $u_1 \in \Re$ and incur a cost $g(x_1, u_1)$, where g is a real-valued function that is bounded either above or below. Thus a cost is incurred only at the second stage.

A policy $\pi = \{\mu_0, \mu_1\}$ is a pair of functions from state to control, i.e., if π is employed and x_0 is the initial state, then $u_0 = \mu_0(x_0)$, and if x_1 is the subsequent state, then $u_1 = \mu_1(x_1)$. The expected value of the cost corresponding to π when x_0 is the initial state is given by

$$J_\pi(x_0) = \int g\big(x_1, \mu_1(x_1)\big)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big). \tag{A.1}$$

We wish to find π to minimize $J_\pi(x_0)$.

To formulate the problem properly, we must insure that the integral in Eq. (A.1) is defined. Various sufficient conditions can be used for this; for example it is sufficient that g, μ_0, and μ_1 be Borel measurable, and that $p(B \mid x_0, u_0)$ is a Borel measurable function of (x_0, u_0) for every Borel set B (see [BeS78]). However, our aim in this example is to discuss the necessary measure theoretic framework not only for the cost $J_\pi(x_0)$ to be defined, but also for the major DP-related results to hold. We thus leave unspecified for the moment the assumptions on the problem data and the measurability restrictions on the policy π.

The optimal cost is

$$J^*(x_0) = \inf_\pi J_\pi(x_0),$$

where the infimum is over all policies $\pi = \{\mu_0, \mu_1\}$ such that μ_0 and μ_1 are measurable functions from $\Re$ to $\Re$ with respect to σ-algebras to be specified later. Given $\epsilon > 0$, a policy π is *ϵ-optimal* if

$$J_\pi(x_0) \le J^*(x_0) + \epsilon, \qquad \forall\ x_0 \in \Re.$$

A policy π is *optimal* if

$$J_\pi(x_0) = J^*(x_0), \qquad \forall\ x_0 \in \Re.$$

The DP Algorithm

The DP algorithm for the preceding two-stage problem takes the form

$$J_1(x_1) = \inf_{u_1 \in \Re} g(x_1, u_1), \qquad \forall\ x_1 \in \Re, \tag{A.2}$$

$$J_0(x_0) = \inf_{u_0 \in \Re} \int J_1(x_1)\, p\big(dx_1 \mid x_0, u_0\big), \qquad \forall\ x_0 \in \Re, \tag{A.3}$$

and assuming that

$$J_0(x_0) > -\infty, \quad \forall\, x_0 \in \Re, \qquad\qquad J_1(x_1) > -\infty, \quad \forall\, x_1 \in \Re,$$

the results we expect to be able to prove are:

R.1: There holds

$$J^*(x_0) = J_0(x_0), \qquad \forall\, x_0 \in \Re.$$

R.2: Given any $\epsilon > 0$, there is an ϵ-optimal policy.

R.3: If $\mu_1^*(x_1)$ and $\mu_0^*(x_0)$ attain the infimum in the DP algorithm (A.2), (A.3) for all $x_1 \in \Re$ and $x_0 \in \Re$, respectively, then $\pi^* = \{\mu_0^*, \mu_1^*\}$ is optimal.

We will see that to establish these results, we will need to address two main issues:

(1) The cost function J_π of a policy π, and the functions J_0 and J_1 produced by DP should be well-defined, with a mathematical framework, which ensures that the integrals in Eqs. (A.1)-(A.3) make sense.

(2) Since $J_0(x_0)$ is easily seen to be a lower bound to $J_\pi(x_0)$ for all x_0 and $\pi = \{\mu_0, \mu_1\}$, the equality of J_0 and J^* will be ensured if the class of policies has an ϵ-selection property, which guarantees that the minima in Eqs. (A.2) and (A.3) can be nearly attained by $\mu_1(x_1)$ and $\mu_0(x_0)$ for all x_1 and x_0, respectively.

To get a better sense of these issues, consider the following informal derivation of R.1:

$$\begin{aligned}
J^*(x_0) &= \inf_{\pi} J_\pi(x_0) \\
&= \inf_{\mu_0} \inf_{\mu_1} \int g\big(x_1, \mu_1(x_1)\big)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big) && \text{(A.4}a\text{)} \\
&= \inf_{\mu_0} \int \left\{ \inf_{\mu_1} g\big(x_1, \mu_1(x_1)\big) \right\} p\big(dx_1 \mid x_0, \mu_0(x_0)\big) && \text{(A.4}b\text{)} \\
&= \inf_{\mu_0} \int \left\{ \inf_{u_1} g(x_1, u_1) \right\} p\big(dx_1 \mid x_0, \mu_0(x_0)\big) \\
&= \inf_{\mu_0} \int J_1(x_1)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big) && \text{(A.4}c\text{)} \\
&= \inf_{u_0} \int J_1(x_1)\, p(dx_1 \mid x_0, u_0) && \text{(A.4}d\text{)} \\
&= J_0(x_0).
\end{aligned}$$

In order to make this derivation meaningful and mathematically rigorous, the following points need to be justified:

(a) g and μ_1 must be such that $g\big(x_1, \mu_1(x_1)\big)$ can be integrated in a well-defined manner in Eq. (A.4a).

(b) The interchange of infimization and integration in Eq. (A.4b) must be legitimate.

(c) g must be such that the function

$$J_1(x_1) = \inf_{u_1} g(x_1, u_1)$$

can be integrated in a well-defined manner in Eq. (A.4c).

We first discuss these points in the easier context where the state space is essentially countable.

Countable Space Problems

We observe that if for each (x_0, u_0), the measure $p(dx_1 \mid x_0, u_0)$ has *countable support*, i.e., is concentrated on a countable number of points, then for a fixed policy π and initial state x_0, the integral defining the cost $J_\pi(x_0)$ of Eq. (A.1) is defined in terms of (possibly infinite) summation. Similarly, the DP algorithm (A.2), (A.3) is defined in terms of summation, and the same is true for the integrals in Eqs. (A.4a)-(A.4d). Thus, there is no need to impose measurability restrictions of any kind for the integrals to make sense, and for the summations/integrations to be well-defined, it is sufficient that g is bounded either above or below.

It can also be shown that the interchange of infimization and summation in Eq. (A.4b) is justified in view of the assumption

$$\inf_{u_1} g(x_1, u_1) > -\infty, \qquad \forall\ x_1 \in \Re.$$

To see this, for any $\epsilon > 0$, select $\bar{\mu}_1 : \Re \mapsto \Re$ such that

$$g\big(x_1, \bar{\mu}_1(x_1)\big) \leq \inf_{u_1} g(x_1, u_1) + \epsilon, \qquad \forall\ x_1 \in \Re. \tag{A.5}$$

Then

$$\begin{aligned}\inf_{\mu_1} \int g\big(x_1, \mu_1(x_1)\big)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big) &\\ &\leq \int g\big(x_1, \bar{\mu}_1(x_1)\big)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big)\\ &\leq \int \inf_{u_1} g(x_1, u_1)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big) + \epsilon.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, it follows that

$$\inf_{\mu_1} \int g\big(x_1, \mu_1(x_1)\big)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big) \leq \int \inf_{u_1} g(x_1, u_1)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big).$$

The reverse inequality also holds, since for all μ_1, we can write

$$\int \inf_{u_1} g(x_1, u_1)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big) \leq \int g\big(x_1, \mu_1(x_1)\big)\, p\big(dx_1 \mid x_0, \mu_0(x_0)\big),$$

and then we can take the infimum over μ_1. It follows that the interchange of infimization and summation in Eq. (A.4b) is justified, with the ϵ-optimal selection property of Eq. (A.5) being the key step in the proof.

We have thus shown that when the measure $p(dx_1 \mid x_0, u_0)$ has countable support, g is bounded either above or below, and $J_0(x_0) > -\infty$ for all x_0 and $J_1(x_1) > -\infty$ for all x_1, the derivation of Eq. (A.4) is valid and proves that the DP algorithm produces the optimal cost function J^* (cf. property R.1).† A similar argument proves the existence of an ϵ-optimal policy (cf. R.2); it uses the ϵ-optimal selection (A.5) for the second stage and a similar ϵ-optimal selection for the first stage, i.e., the existence of a $\bar{\mu}_0 : \Re \mapsto \Re$ such that

$$\int J_1(x_1)\, p\big(dx_1 \mid x_0, \bar{\mu}_0(x_0)\big) \leq \inf_{u_0} \int J_1(x_1)\, p(dx_1 \mid x_0, u_0) + \epsilon. \quad \text{(A.6)}$$

Also R.3 follows easily using the fact that there are no measurability restrictions on μ_0 and μ_1.

Approaches for Uncountable Space Problems

To address the case where $p(dx_1 \mid x_0, u_0)$ does not have countable support, two approaches have been used. The first is to *expand the notion of integration*, and the second is to place *appropriate measurability restrictions on g, p, and* $\{\mu_0, \mu_1\}$. Expanding the notion of integration is possible by interpreting the integrals appearing in the preceding equations as outer integrals. Since the outer integral can be defined for any function, measurable or not, there is no need to impose any measurability assumptions, and the arguments given above go through just as in the countable disturbance case. We do not discuss this approach further except to mention that the Bertsekas and Shreve book [BeS78] shows that the basic results for finite and infinite horizon problems of perfect state information carry through within an outer integration framework. However, there are inherent limitations in this approach centering around the pathologies of outer integration, as discussed in [BeS78].

The second approach is to impose a suitable measurability structure that allows the key proof steps of the validity of the DP algorithm. These are:

† The condition that g is bounded either above or below may be replaced by any condition that guarantees that the infinite sum/integral of J_1 in Eq. (A.3) is well-defined. Note also that if g is bounded below, then the assumption that $J_0(x_0) > -\infty$ for all x_0 and $J_1(x_1) > -\infty$ for all x_1 is automatically satisfied.

(a) Properly interpreting the integrals in the definition (A.2)-(A.3) of the DP algorithm and the derivation (A.4).

(b) The ϵ-optimal selection property (A.5), which in turn justifies the interchange of infimization and integration in Eq. (A.4b).

To enable (a), the required properties of the problem structure must include the preservation of measurability under partial minimization. In particular, it is necessary that when g is measurable in some sense, the partial minimum function

$$J_1(x_1) = \inf_{u_1} g(x_1, u_1)$$

is also measurable in the same sense, so that the integration in Eq. (A.3) is well-defined. It turns out that this is a major difficulty with Borel measurability, which may appear to be a natural framework for formulating the problem: *J_1 need not be Borel measurable even when g is Borel measurable*. For this reason it is necessary to pass to a larger class of measurable functions, which is closed under the key operation of partial minimization (and also under some other common operations, such as addition and functional composition).†

One such class is *lower semianalytic functions* and the related class of *universally measurable functions*, which will be the focus of the next section. They are the basis for a problem formulation that enables a DP theory as powerful as the one for problems where measurability is of no concern (e.g., those where the state and control spaces are countable).

A.2 RESOLUTION OF THE MEASURABILITY ISSUES

The example of the preceding section indicates that if measurability restrictions are necessary for the problem data and policies, then measurable selection and preservation of measurability under partial minimization, become crucial parts of the analysis. We will discuss measurability frameworks that are favorable in this regard, and to this end, we will use the theory of Borel spaces.

† It is also possible to use a smaller class of functions that is closed under the same operations. This has led to the so-called *semicontinuous models*, where the state and control spaces are Borel spaces, and g and p have certain semicontinuity and other properties. These models are also analyzed in detail in the Bertsekas and Shreve book [BeS78] (Section 8.3). However, they are not as useful and widely applicable as the universally measurable models we will focus on, because they involve assumptions that may be restrictive and/or hard to verify. By contrast, the universally measurable models are simple and very general. They allow a problem formulation that brings to bear the power of DP analysis under minimal assumptions. This analysis can in turn be used to prove more specific results based on special characteristics of the model.

Borel Spaces and Analytic Sets

Given a topological space Y, we denote by $\mathcal{B}_Y$ the σ-algebra generated by the open subsets of Y, and refer to the members of $\mathcal{B}_Y$ as the *Borel subsets* of Y. A topological space Y is a *Borel space* if it is homeomorphic to a Borel subset of a complete separable metric space. The concept of Borel space is quite broad, containing any "reasonable" subset of n-dimensional Euclidean space. Any Borel subset of a Borel space is again a Borel space, as is any homeomorphic image of a Borel space and any finite or countable Cartesian product of Borel spaces. Let Y and Z be Borel spaces, and consider a function $h : Y \mapsto Z$. We say that h is *Borel measurable* if $h^{-1}(B) \in \mathcal{B}_Y$ for every $B \in \mathcal{B}_Z$.

Borel spaces have a deficiency in the context of optimization: even in the unit square, there exist Borel sets whose projections onto an axis are not Borel subsets of that axis. In fact, this is the source of the difficulty we mentioned earlier regarding Borel measurability in the DP context: if $g(x_1, u_1)$ is Borel measurable, the partial minimum function

$$J_1(x_1) = \inf_{u_1} g(x_1, u_1)$$

need not be, because its level sets are defined in terms of projections of the level sets of g as

$$\big\{x_1 \mid J_1(x_1) < c\big\} = P\Big(\big\{(x_1, u_1) \mid g(x_1, u_1) < c\big\}\Big),$$

where c is a scalar and $P(\cdot)$ denotes projection on the space of x_1. As an example, take g to be the indicator of a Borel subset of the unit square whose projection on the x_1-axis is not Borel. Then J_1 is the indicator function of this projection, so it is not Borel measurable. This leads us to the notion of an analytic set.

A subset A of a Borel space Y is said to be *analytic* if there exists a Borel space Z and a Borel subset B of $Y \times Z$ such that $A = \text{proj}_Y(B)$, where proj_Y is the projection mapping from $Y \times Z$ to Y. It is clear that every Borel subset of a Borel space is analytic.

Analytic sets have many interesting properties, which are discussed in detail in [BeS78]. Some of these properties are particularly relevant to DP analysis. For example, let Y and Z be Borel spaces. Then:

(i) If $A \subset Y$ is analytic and $h : Y \mapsto Z$ is Borel measurable, then $h(A)$ is analytic. In particular, if Y is a product of Borel spaces Y_1 and Y_2, and $A \subset Y_1 \times Y_2$ is analytic, then $\text{proj}_{Y_1}(A)$ is analytic. Thus, the class of analytic sets is closed with respect to projection, a critical property for DP, which the class of Borel sets is lacking, as mentioned earlier.

(ii) If $A \subset Z$ is analytic and $h : Y \mapsto Z$ is Borel measurable, then $h^{-1}(A)$ is analytic.

(iii) If $A_1, A_2, \ldots$ are analytic subsets of Y, then $\cup_{k=1}^{\infty} A_k$ and $\cap_{k=1}^{\infty} A_k$ are analytic.

However, the complement of an analytic set need not be analytic, so the collection of analytic subsets of Y need not be a σ-algebra.

Lower Semianalytic Functions

Let Y be a Borel space and let $h : Y \mapsto [-\infty, \infty]$ be a function. We say that h is *lower semianalytic* if the level set

$$\{y \in Y \mid h(y) < c\}$$

is analytic for every $c \in \Re$. The following proposition states that lower analyticity is preserved under partial minimization, a key result for our purposes. The proof follows from the preservation of analyticity of a subset of a product space under projection onto one of the component spaces, as in (i) above (see [BeS78], Prop. 7.47).

Proposition A.1: Let Y and Z be Borel spaces, and let $h : Y \times Z \mapsto [-\infty, \infty]$ be lower semianalytic. Then $h^* : Y \mapsto [-\infty, \infty]$ defined by

$$h^*(y) = \inf_{z \in Z} h(y, z)$$

is lower semianalytic.

By comparing the DP equation $J_1(x_1) = \inf_{u_1} g(x_1, u_1)$ [cf. Eq. (A.2)] and Prop. A.1, we see how lower semianalytic functions can arise in DP. In particular, J_1 is lower semianalytic if g is. Let us also give two additional properties of lower semianalytic functions that play an important role in DP (for a proof, see [BeS78], Lemma 7.40).

Proposition A.2: Let Y be a Borel space, and let $h : Y \mapsto [-\infty, \infty]$ and $l : Y \mapsto [-\infty, \infty]$ be lower semianalytic. Suppose that for every $y \in Y$, the sum $h(y) + l(y)$ is defined, i.e., is not of the form $\infty - \infty$. Then $h + l$ is lower semianalytic.

Proposition A.3: Let Y and Z be Borel spaces, let $h : Y \mapsto Z$ be Borel measurable, and let $l : Z \mapsto [-\infty, \infty]$ be lower semianalytic. Then the composition $l \circ h$ is lower semianalytic.

Universal Measurability

To address questions relating to the definition of the integrals appearing in the DP algorithm, we must discuss the measurability properties of lower semianalytic functions. In addition to the Borel σ-algebra $\mathcal{B}_Y$ mentioned earlier, there is the *universal σ-algebra* $\mathcal{U}_Y$, which is the intersection of all completions of $\mathcal{B}_Y$ with respect to all probability measures. Thus, $E \in \mathcal{U}_Y$ if and only if, given any probability measure p on $(Y, \mathcal{B}_Y)$, there is a Borel set B and a p-null set N such that $E = B \cup N$. Clearly, we have $\mathcal{B}_Y \subset \mathcal{U}_Y$. It is also true that every analytic set is universally measurable (for a proof, see [BeS78], Corollary 7.42.1), and hence the σ-algebra generated by the analytic sets, called the *analytic σ-algebra*, and denoted $\mathcal{A}_Y$, is contained in $\mathcal{U}_Y$:

$$\mathcal{B}_Y \subset \mathcal{A}_Y \subset \mathcal{U}_Y.$$

Let X, Y, and Z be Borel spaces, and consider a function $h : Y \mapsto Z$. We say that h is *universally measurable* if $h^{-1}(B) \in \mathcal{U}_Y$ for every $B \in \mathcal{B}_Z$. It can be shown that if $U \subset Z$ is universally measurable and h is universally measurable, then $h^{-1}(U)$ is also universally measurable. As a result, if $g : X \mapsto Y$, $h : Y \mapsto Z$ are universally measurable functions, then the composition $(g \circ h) : X \mapsto Z$ is universally measurable.

We say that $h : Y \mapsto Z$ is *analytically measurable* if $h^{-1}(B) \in \mathcal{A}_Y$ for every $B \in \mathcal{B}_Z$. It can be seen that *every lower semianalytic function is analytically measurable*, and in view of the inclusion $\mathcal{A}_Y \subset \mathcal{U}_Y$, it is *also universally measurable*.

Integration of Lower Semianalytic Functions

If p is a probability measure on $(Y, \mathcal{B}_Y)$, then p has a unique extension to a probability measure $\bar{p}$ on $(Y, \mathcal{U}_Y)$. We write simply p instead of $\bar{p}$ and $\int h dp$ in place of $\int h d\bar{p}$. In particular, if h is lower semianalytic, then $\int h\, dp$ is interpreted in this manner.

Let Y and Z be Borel spaces. A *stochastic kernel* $q(dz \mid y)$ on Z given Y is a collection of probability measures on $(Z, \mathcal{B}_Z)$ parameterized by the elements of Y. If for each Borel set $B \in \mathcal{B}_Z$, the function $q(B \mid y)$ is Borel measurable (universally measurable) in y, the stochastic kernel $q(dz \mid y)$ is said to be *Borel measurable* (*universally measurable*, respectively). The following proposition provides another basic property for the DP context (for a proof, see [BeS78], Props. 7.46 and 7.48).

Proposition A.4: Let Y and Z be Borel spaces, and let $q(dz \mid y)$ be a stochastic kernel on Z given Y. Let also $h : Y \times Z \mapsto [-\infty, \infty]$ be a function that is bounded either above or below.

(a) If q is Borel measurable and h is lower semianalytic, then the function $l : Y \mapsto [-\infty, \infty]$ given by

$$l(y) = \int_Z h(y, z)q(dz \mid y)$$

is lower semianalytic.

(b) If q is universally measurable and h is universally measurable, then the function $l : Y \mapsto [-\infty, \infty]$ given by

$$l(y) = \int_Z h(y, z)q(dz \mid y)$$

is universally measurable.

Note that the boundedness above or below assumption on h in the preceding proposition aims to ensure that $l(y)$ is well-defined for every y as an integral.†

Returning to the DP algorithm (A.2)-(A.3) of Section A.1, note that if the cost function g is lower semianalytic and bounded either above or below, then the partial minimum function J_1 given by the DP Eq. (A.2) is lower semianalytic (cf. Prop. A.1), and bounded either above or below, respectively. Furthermore, if the transition kernel $p(dx_1 \mid x_0, u_0)$ is Borel measurable, then the integral

$$\int J_1(x_1)\, p\big(dx_1 \mid x_0, u_0\big) \tag{A.7}$$

is a lower semianalytic function of (x_0, u_0) (cf. Prop. A.4), and in view of Prop. A.1, the same is true of the function J_0 given by the DP Eq. (A.3), which is the partial minimum over u_0 of the expression (A.7). Thus, with

† We use here the classical definition of integral, whereby for a probability measure p, the integral of an extended real-valued function f, with positive and negative parts f^+ and f^-, is defined as

$$\int f dp = \int f^+ dp - \int f^- dp,$$

provided $\int f^+ dp < \infty$ or $\int f^- dp < \infty$. The book [BeS78] (Section 7.4.4) uses a more general definition, which adopts the rule $\infty - \infty = \infty$ for the case where $\int f^+ dp = \infty$ and $\int f^- dp = \infty$. With this expanded definition of integral, there is no need for the boundedness assumption in Prop. A.4 (cf. [BeS78], Props. 7.46 and 7.48).

lower semianalytic g and Borel measurable p, the integrals appearing in the DP algorithm make sense.

Note that in the example of Section A.1, there is no cost incurred in the first stage of the system operation. When such a cost, call it $g_0(x_0, u_0)$, is introduced, the expression minimized in the DP Eq. (A.3) becomes

$$g_0(x_0, u_0) + \int J_1(x_1)\, p\big(dx_1 \mid x_0, u_0\big),$$

which is still a lower semianalytic function of (x_0, u_0), provided g_0 is lower semianalytic and the sum above is not of the form $\infty - \infty$ for any (x_0, u_0) (Prop. A.2). Also, for alternative models defined in terms of a system function rather than a stochastic kernel (e.g., the total cost model of Chapter 1), Prop. A.3 provides some of the necessary machinery to show that the functions generated by the DP algorithm are lower semianalytic.

Universally Measurable Selection

The preceding discussion has shown that if g is lower semianalytic and bounded either above or below, and p is Borel measurable, the DP algorithm (A.2)-(A.3) is well-defined and produces lower semianalytic functions J_1 and J_0. However, this does not by itself imply that J_0 is equal to the optimal cost function J^*. For this it is necessary that the chosen class of policies has the ϵ-optimal selection property (A.5). It turns out that universally measurable policies have this property.

The following is the key selection theorem given in a general form, which also addresses the question of existence of optimal policies that can be obtained from the DP algorithm (for a proof, see [BeS78], Prop. 7.50). The theorem shows that if any functions $\bar{\mu}_1 : \Re \to \Re$ and $\bar{\mu}_0 : \Re \to \Re$ can be found such that $\bar{\mu}_1(x_1)$ and $\bar{\mu}_0(x_0)$ attain the respective minima in Eqs. (A.2) and (A.3), for every x_1 and x_0, then $\bar{\mu}_1$ and $\bar{\mu}_0$ can be chosen to be universally measurable, the DP algorithm yields the optimal cost function and $\pi = (\bar{\mu}_0, \bar{\mu}_1)$ is optimal, provided that g is lower semianalytic and the integral in Eq. (A.3) is a lower semianalytic function of (x_0, u_0).

Proposition A.5: (Measurable Selection Theorem) Let Y and Z be Borel spaces and let $h : Y \times Z \mapsto [-\infty, \infty]$ be lower semianalytic. Define $h^* : Y \mapsto [-\infty, \infty]$ by

$$h^*(y) = \inf_{z \in Z} h(y, z),$$

and let

$$I = \big\{ y \in Y \mid \text{there exists a } z_y \in Z \text{ for which } h(y, z_y) = h^*(y) \big\},$$

i.e., I is the set of points y for which the infimum above is attained. For any $\epsilon > 0$, there exists a universally measurable function $\phi : Y \mapsto Z$ such that

$$h\big(y, \phi(y)\big) = h^*(y), \qquad \forall\ y \in I,$$

$$h\big(y, \phi(y)\big) \leq \begin{cases} h^*(y) + \epsilon, & \forall\ y \notin I \text{ with } h^*(y) > -\infty, \\ -1/\epsilon, & \forall\ y \notin I \text{ with } h^*(y) = -\infty. \end{cases}$$

Universal Measurability Framework: A Summary

In conclusion, the preceding discussion shows that in the two-stage example of Section A.1, the measurability issues are resolved in the following sense: the DP algorithm (A.2)-(A.3) is well-defined, produces lower semianalytic functions J_1 and J_0, and yields the optimal cost function (as in R.1), and furthermore there exist ϵ-optimal and possibly exactly optimal policies (as in R.2 and R.3), provided that:

(a) *The stage cost function g is lower semianalytic and is bounded either above or below.* Lower analyticity is needed to show that the function J_1 of the DP Eq. (A.2) is lower semianalytic and hence also universally measurable (cf. Prop. A.1). Boundedness either above or below is needed to ensure the respective boundedness property for J_1, which will be needed to guarantee that the integral of J_1 in Eq. (A.3) is defined (according to the classical definition). The more "natural" Borel measurability assumption on g implies lower analyticity of g, but will not keep the functions J_1 and J_0 produced by the DP algorithm within the domain of Borel measurability. This is because the partial minimum operation on Borel measurable functions takes us outside that domain (cf. Prop. A.1).

(b) *The stochastic kernel p is Borel measurable.* This is needed in order for the integral in the DP Eq. (A.3) to be defined as a lower semianalytic function of (x_0, u_0) (cf. Prop. A.4). In turn, this is used to show that the function J_0 of the DP Eq. (A.3) is lower semianalytic (cf. Prop. A.1).

(c) *The control functions μ_0 and μ_1 are allowed to be universally measurable, and we have $J_0(x_0) > -\infty$ for all x_0 and $J_1(x_1) > -\infty$ for all x_1.* This is needed in order for the calculation of Eq. (A.4) to go through (using the measurable selection property of Prop. A.5), and

show that the DP algorithm produces the optimal cost function (cf. R.1). It is also needed (using again Prop. A.5) in order to show the associated existence of solutions results (cf. R.2 and R.3).

Extension to General Finite-Horizon DP

Let us now extend our analysis to an N-stage model with state x_k and control u_k that take values in Borel spaces X and U, respectively. We assume stochastic/transition kernels $p_k(dx_{k+1} \mid x_k, u_k)$, which are Borel measurable, and stage cost functions $g_k : X \times U \mapsto (-\infty, \infty]$, which are lower semianalytic and bounded either above or below.† Furthermore, we allow policies $\pi = \{\mu_0, \ldots, \mu_{N-1}\}$ that are randomized: each component μ_k is a universally measurable stochastic kernel $\mu_k(du_k \mid x_k)$ from X to U. If for every x_k and k, $\mu_k(du_k \mid x_k)$ assigns probability 1 to a single control u_k, π is said to be *nonrandomized*.

Each policy π and initial state x_0 define a unique probability measure with respect to which $g_k(x_k, u_k)$ can be integrated to produce the expected value of g_k. The sum of these expected values for $k = 0, \ldots, N-1$, is the cost $J_\pi(x_0)$. It is convenient to write this cost in terms of the following DP-like backwards recursion (see [BeS78], Section 8.1):

$$J_{\pi,N-1}(x_{N-1}) = \int g_{N-1}(x_{N-1}, u_{N-1}) \mu_{N-1}(du_{N-1} \mid x_{N-1}),$$

$$J_{\pi,k}(x_k) = \int \left(g_k(x_k, u_k) + \int J_{\pi,k+1}(x_{k+1})\, p_k(dx_{k+1} \mid x_k, u_k) \right) \mu_k(du_k \mid x_k), \qquad k = 0, \ldots, N-2.$$

The function obtained at the last step is the cost of π starting at x_0:

$$J_\pi(x_0) = J_{\pi,0}(x_0).$$

We can interpret $J_{\pi,k}(x_k)$ as the expected cost-to-go starting from x_k at time k, and using π. Note that by Prop. A.4, the functions $J_{\pi,k}$ are all universally measurable.

The DP algorithm is given by

$$J_{N-1}(x_{N-1}) = \inf_{u_{N-1} \in U} g_{N-1}(x_{N-1}, u_{N-1}), \qquad \forall\ x_{N-1},$$

$$J_k(x_k) = \inf_{u_k \in U} \left[g_k(x_k, u_k) + \int J_{k+1}(x_{k+1})\, p_k\big(dx_{k+1} \mid x_k, u_k\big) \right], \quad \forall\ x_k,\ k.$$

† Note that since g_k may take the value ∞, constraints of the form $u_k \in U_k(x_k)$ may be implicitly introduced by letting $g_k(x_k, u_k) = \infty$ when $u_k \notin U_k(x_k)$.

By essentially replicating the analysis of the two-stage example, we can show that the integrals in the above DP algorithm are well-defined, and that the functions $J_{N-1}, \ldots, J_0$ are lower semianalytic.

It can be seen from the preceding expressions that we have for all policies π

$$J_k(x_k) \leq J_{\pi,k}(x_k), \qquad \forall\ x_k,\ k = 0, \ldots, N-1.$$

To show equality within $\epsilon \geq 0$ in the above relation, we may use the measurable selection theorem (Prop. A.5), assuming that

$$J_k(x_k) > -\infty, \quad \forall\ x_k,\ k,$$

so that ϵ-optimal universally measurable selection is possible in the DP algorithm. In particular, define $\overline{\pi} = \{\overline{\mu}_0, \ldots, \overline{\mu}_{N-1}\}$ such that $\overline{\mu}_k : X \mapsto U$ is universally measurable, and for all x_k and k,

$$g_k\big(x_k, \overline{\mu}_k(x_k)\big) + \int J_{k+1}(x_{k+1})\, p_k\big(dx_{k+1} \mid x_k, \overline{\mu}_k(x_k)\big) \leq J_k(x_k) + \frac{\epsilon}{N}. \tag{A.8}$$

Then, we can show by induction that

$$J_k(x_k) \leq J_{\overline{\pi},k}(x_k) \leq J_k(x_k) + \frac{(N-k)\epsilon}{N}, \qquad \forall\ x_k,\ k = 0, \ldots, N-1,$$

and in particular, for $k = 0$,

$$J_0(x_0) \leq J_{\overline{\pi}}(x_0) \leq J_0(x_0) + \epsilon, \qquad \forall\ x_0.$$

and hence also

$$J^*(x_0) = \inf_{\pi} J_\pi(x_0) = J_0(x_0).$$

Thus, the DP algorithm produces the optimal cost function, and via the approximate minimization of Eq. (A.8), an ϵ-optimal policy. Similarly, if the infimum is attained for all x_k and k in the DP algorithm, then there exists an optimal policy. Note that both the ϵ-optimal and the exact optimal policies can be taken be nonrandomized.

An interesting characteristic of the preceding line of development is that it decouples the issue of the definition of the DP algorithm from the question of whether it yields the optimal cost function and ϵ-optimal or nearly optimal policies. In the former question, the key fact is the preservation of lower semianalyticity under partial minimization and integration, while in the latter question, the key fact is whether ϵ-optimal selection is possible in the DP algorithm within the class of policies stipulated. To illustrate this point, suppose that we are interested in optimizing the cost $J_\pi(x_0)$ over a *restricted subset* Π of the randomized universally measurable policies. For example in problems with special structure, Π may be a class

of continuous functions, or linear functions, or functions with some special structural characteristics [e.g., (s, S) or other threshold policies in inventory control]. Then, Borel measurability of the stochastic kernels and lower semianalyticity of the costs per stage will guarantee that the functions J_k produced by the DP algorithm are well-defined and can be analyzed. If the analysis shows that the class of policies Π has the ϵ-selection property (A.8), then it follows that $J_0(x_0)$ is equal to the optimal cost over the restricted class Π, and that ϵ-optimal policies exist within this class.

The assumptions of Borel measurability of the stochastic kernels, lower semianalyticity of the costs per stage, and universally measurable policies, are the basis for the framework adopted by Bertsekas and Shreve [BeS78], which provides a comprehensive analysis of finite and infinite horizon total cost problems. The results obtained there using this framework closely parallel the results of Chapters 1 and 3 of the present volume, but apply to the more general case of uncountable disturbance spaces. There is also additional analysis in [BeS78] on problems of imperfect state information, as well as various refinements of the measurability framework just described. Among others, these refinements involve analytically measurable policies, and limit measurable policies (measurable with respect to the, so-called, limit σ-algebra, the smallest σ-algebra that has the properties necessary for a DP theory that is comparably powerful to the one for the universal σ-algebra).

Issues of Policy Iteration

A difficulty with the universally measurable framework, which was left unresolved in [BeS78], relates to the standard policy iteration method. The issue is that the selection of an admissible measurable policy can fail at the policy improvement step because the cost function of an analytically or universally measurable policy, produced by policy evaluation, need not have the lower semianalytic structure for exact or ϵ-exact selection of an improved policy, which is required by Prop. A.5. This causes the policy iteration procedure to break down.

The recent paper by Yu and Bertsekas [YuB15] provides an algorithmic approach to circumvent this difficulty, and to allow stationary policies to be used in computing the optimal cost function, in a manner that resembles policy iteration (even when ϵ-optimal stationary policies do not exist). The approach is based on an algorithm that combines characteristics of both value and policy iteration.

Algorithmically, compared to standard policy iteration, the main difference of this method is in the policy evaluation phase: instead of computing the cost function of a given policy, it solves exactly or approximately an optimal stopping problem defined by a stationary policy of interest and by a stopping cost that is an estimate of the optimal cost. The stopping costs are then adjusted and the procedure is repeated. This is a similar idea to

the one used to construct a policy iteration method with a uniform fixed point (Sections 2.6.3 and 3.5.4), but it serves here the different purpose of circumventing the measure-theoretic difficulties.

To avoid measurability issues, the fact that every universally measurable stationary policy has Borel measurable portions is used, and the optimal stopping problems are defined accordingly so that the iterative method just mentioned can operate within the family of functions with the desired semianalytic structure. We refer to [YuB15] for the details of the analysis.

Let us finally note that the paper [YuB15] contains several new results for the undiscounted problems of Section 4.1 under Assumption P and N (but always within a measure-theoretic framework). Noteworthy in this regard is a convergence result for value iteration under Assumption P, starting from initial conditions J satisfying $J^* \leq J \leq cJ^*$ for some scalar $c > 1$ (Section 5.1 of [YuB15]).

[BeS78] Bertsekas, D. P., and Shreve, S. E., 1978. Stochastic Optimal Control: The Discrete Time Case, Academic Press, N. Y.; may be downloaded from http://web.mit.edu/dimitrib/www/home.html

[YuB15] Yu, H., and Bertsekas, D. P., 2015. "A Mixed Value and Policy Iteration Method for Stochastic Control with Universally Measurable Policies," Math. of OR, Vol. 40, pp. 926-968.