*Chapter 5*

# Infinite Horizon Models under Monotonicity Assumptions

## 5.1 General Remarks and Assumptions

Consider the infinite horizon problem

$$\begin{aligned} &\text{minimize} \quad J_\pi(x) = \lim_{N\to\infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0)(x) \\ &\text{subject to} \quad \pi = (\mu_0, \mu_1, \ldots) \in \Pi. \end{aligned} \tag{1}$$

In this chapter we impose monotonicity assumptions on the function $J_0$ which guarantee that $J_\pi$ is well defined for all $\pi \in \Pi$. For every result to be shown in this chapter, one of the following two assumptions will be in effect.

**Assumption I** (Uniform Increase Assumption) There holds

$$J_0(x) \le H(x, u, J_0) \qquad \forall x \in S, \quad u \in U(x). \tag{2}$$

**Assumption D** (Uniform Decrease Assumption) There holds

$$J_0(x) \ge H(x, u, J_0) \qquad \forall x \in S, \quad u \in U(x). \tag{3}$$

It is easy to see that under each of these assumptions the limit in (1) is well defined as a real number or $\pm\infty$. Indeed, in the case of Assumption I we have from (2) that

$$J_0 \leq T_{\mu_0}(J_0) \leq (T_{\mu_0} T_{\mu_1})(J_0) \leq \cdots \leq (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0) \leq \cdots,$$

while in the case of Assumption D we have from (3) that

$$J_0 \geq T_{\mu_0}(J_0) \geq (T_{\mu_0} T_{\mu_1})(J_0) \geq \cdots \geq (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0) \geq \cdots.$$

In both cases, the limit in (1) clearly exists in the extended real numbers for each $x \in S$.

In our analysis under Assumptions I or D we will occasionally need to assume one or more of the following continuity properties for the mapping $H$. Assumptions I.1 and I.2 will be used in conjunction with Assumption I, while Assumptions D.1 and D.2 will be used in conjunction with Assumption D.

**Assumption I.1** If $\{J_k\} \subset F$ is a sequence satisfying $J_0 \leq J_k \leq J_{k+1}$ for all $k$, then

$$\lim_{k\to\infty} H(x, u, J_k) = H\left(x, u, \lim_{k\to\infty} J_k\right) \qquad \forall x \in S, \quad u \in U(x). \tag{4}$$

**Assumption I.2** There exists a scalar $\alpha > 0$ such that for all scalars $r > 0$ and functions $J \in F$ with $J_0 \leq J$, we have

$$H(x, u, J) \leq H(x, u, J + r) \leq H(x, u, J) + \alpha r \qquad \forall x \in S, \quad u \in U(x). \tag{5}$$

**Assumption D.1** If $\{J_k\} \subset F$ is a sequence satisfying $J_{k+1} \leq J_k \leq J_0$ for all $k$, then

$$\lim_{k\to\infty} H(x, u, J_k) = H\left(x, u, \lim_{k\to\infty} J_k\right) \qquad \forall x \in S, \quad u \in U(x). \tag{6}$$

**Assumption D.2** There exists a scalar $\alpha > 0$ such that for all scalars $r > 0$ and functions $J \in F$ with $J \leq J_0$, we have

$$H(x, u, J) - \alpha r \leq H(x, u, J - r) \leq H(x, u, J) \qquad \forall x \in S, \quad u \in U(x). \tag{7}$$

## 5.2 The Optimality Equation

We first consider the question whether the optimality equation $J^* = T(J^*)$ holds. As a preliminary step we prove the following result, which is of independent interest.

**Proposition 5.1** Let Assumptions I, I.1, and I.2 hold. Then given any $\varepsilon > 0$, there exists an $\varepsilon$-optimal policy, i.e., a $\pi_\varepsilon \in \Pi$, such that

$$J^* \leq J_{\pi_\varepsilon} \leq J^* + \varepsilon. \tag{8}$$

Furthermore, if the scalar $\alpha$ in I.2 satisfies $\alpha < 1$, the policy $\pi_\varepsilon$ can be taken to be stationary.

*Proof* Let $\{\varepsilon_k\}$ be a sequence such that $\varepsilon_k > 0$ for all $k$ and

$$\sum_{k=0}^{\infty} \alpha^k \varepsilon_k = \varepsilon. \tag{9}$$

For each $x \in S$, consider a sequence of policies $\{\pi_k[x]\} \subset \Pi$ of the form

$$\pi_k[x] = (\mu_0^k[x], \mu_1^k[x], \ldots),$$

such that for $k = 0, 1, \ldots$

$$J_{\pi_k[x]}(x) \le J^*(x) + \varepsilon_k \qquad \forall x \in S. \tag{10}$$

Such a sequence exists, since we have $J^*(x) > -\infty$ under our assumptions.

The (admittedly confusing) notation used here and later in the proof should be interpreted as follows. The policy $\pi_k[x] = (\mu_0^k[x], \mu_1^k[x], \ldots)$ is associated with $x$. Thus $\mu_i^k[x]$ denotes, for each $x \in S$ and $k$, a function in $M$, while $\mu_i^k[x](z)$ denotes the value of $\mu_i^k[x]$ at an element $z \in S$. In particular, $\mu_i^k[x](x)$ denotes the value of $\mu_i^k[x]$ at $x$.

Consider the functions $\bar{\mu}_k \in M$ defined by

$$\bar{\mu}_k(x) = \mu_0^k[x](x) \qquad \forall x \in S \tag{11}$$

and the functions $\bar{J}_k$ defined by

$$\bar{J}_k(x) = H\left[x, \bar{\mu}_k(x), \lim_{i \to \infty} (T_{\mu_1^k[x]} \cdots T_{\mu_i^k[x]})(J_0)\right] \qquad \forall x \in S, \quad k = 0, 1, \ldots. \tag{12}$$

By using (10), (11), I, and I.1, we obtain

$$\begin{aligned} \bar{J}_k(x) &= \lim_{i \to \infty} (T_{\mu_0^k[x]} \cdots T_{\mu_i^k[x]})(J_0)(x) \\ &= J_{\pi_k[x]}(x) \le J^*(x) + \varepsilon_k, \qquad \forall x \in S, \quad k = 0, 1, \ldots. \end{aligned} \tag{13}$$

We have from (12), (13), and I.2 for all $k = 1, 2, \ldots$ and $x \in S$

$$\begin{aligned} T_{\bar{\mu}_{k-1}}(\bar{J}_k)(x) &= H[x, \bar{\mu}_{k-1}(x), \bar{J}_k] \\ &\le H[x, \bar{\mu}_{k-1}(x), (J^* + \varepsilon_k)] \\ &\le H[x, \bar{\mu}_{k-1}(x), J^*] + \alpha\varepsilon_k \\ &\le H[x, \bar{\mu}_{k-1}(x), \lim_{i \to \infty} (T_{\mu_1^{k-1}[x]} \cdots T_{\mu_i^{k-1}[x]})(J_0)] + \alpha\varepsilon_k \\ &= \bar{J}_{k-1}(x) + \alpha\varepsilon_k, \end{aligned}$$

and finally,

$$T_{\bar{\mu}_{k-1}}(\bar{J}_k) \le \bar{J}_{k-1} + \alpha\varepsilon_k, \qquad k = 1, 2, \ldots. \tag{14}$$

Using this inequality and I.2, we obtain

$$\begin{aligned} T_{\bar{\mu}_{k-2}}[T_{\bar{\mu}_{k-1}}(\bar{J}_k)] &\leq T_{\bar{\mu}_{k-2}}(\bar{J}_{k-1} + \alpha\varepsilon_k) \\ &\leq T_{\bar{\mu}_{k-2}}(\bar{J}_{k-1}) + \alpha^2\varepsilon_k \leq \bar{J}_{k-2} + (\alpha\varepsilon_{k-1} + \alpha^2\varepsilon_k). \end{aligned}$$

Continuing in the same manner, we obtain for $k = 1, 2, \ldots$

$$(T_{\bar{\mu}_0} \cdots T_{\bar{\mu}_{k-1}})(\bar{J}_k) \leq \bar{J}_0 + (\alpha\varepsilon_1 + \cdots + \alpha^k\varepsilon_k) \leq J^* + \left(\sum_{i=0}^{k} \alpha^i \varepsilon_i\right).$$

Since $J_0 \leq \bar{J}_k$, it follows that

$$(T_{\bar{\mu}_0} \cdots T_{\bar{\mu}_{k-1}})(J_0) \leq J^* + \left(\sum_{i=0}^{k} \alpha^i \varepsilon_i\right).$$

Denote $\pi_\varepsilon = (\bar{\mu}_0, \bar{\mu}_1, \ldots)$. Then by taking the limit in the preceding inequality and using (9), we obtain

$$J_{\pi_\varepsilon} \leq J^* + \varepsilon.$$

If $\alpha < 1$, we take $\varepsilon_k = \varepsilon(1 - \alpha)$ for all $k$ and $\pi_k[x] = (\mu_0[x], \mu_1[x], \ldots)$ in (10). The stationary policy $\pi_\varepsilon = (\bar{\mu}, \bar{\mu}, \ldots)$, where $\bar{\mu}(x) = \mu_0[x](x)$ for all $x \in S$, satisfies $J_{\pi_\varepsilon} \leq J^* + \varepsilon$. Q.E.D.

It is easy to see that the assumption $\alpha < 1$ is essential in order to be able to take $\pi_\varepsilon$ stationary in the preceding proposition. As an example, take $S = \{0\}$, $U(0) = (0, \infty)$, $J_0(0) = 0$, $H(0, u, J) = u + J(0)$. Then $J^*(0) = 0$, but for any $\mu \in M$, we have $J_\mu(0) = \infty$.

By using Proposition 5.1 we can prove the optimality equation under I, I.1, and I.2.

**Proposition 5.2** Let I, I.1, and I.2 hold. Then

$$J^* = T(J^*).$$

Furthermore, if $J' \in F$ is such that $J' \geq J_0$ and $J' \geq T(J')$, then $J' \geq J^*$.

*Proof* For every $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$ and $x \in S$, we have, from I.1,

$$\begin{aligned} J_\pi(x) &= \lim_{k\to\infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J_0)(x) \\ &= T_{\mu_0}\left[\lim_{k\to\infty} (T_{\mu_1} \cdots T_{\mu_k})(J_0)\right](x) \geq T_{\mu_0}(J^*)(x) \geq T(J^*)(x). \end{aligned}$$

By taking the infimum of the left-hand side over $\pi \in \Pi$, we obtain

$$J^* \geq T(J^*).$$

To prove the reverse inequality, let $\varepsilon_1$ and $\varepsilon_2$ be any positive scalars and let $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots)$ be such that

$$T_{\bar{\mu}_0}(J^*) \leq T(J^*) + \varepsilon_1, \qquad J_{\bar{\pi}_1} \leq J^* + \varepsilon_2,$$

where $\bar{\pi}_1 = (\bar{\mu}_1, \bar{\mu}_2, \ldots)$. Such a policy exists by Proposition 5.1. We have

$$\begin{aligned} J_{\bar{\pi}} &= \lim_{k\to\infty} (T_{\bar{\mu}_0} T_{\bar{\mu}_1} \cdots T_{\bar{\mu}_k})(J_0) \\ &= T_{\bar{\mu}_0}\left[\lim_{k\to\infty} (T_{\bar{\mu}_1} \cdots T_{\bar{\mu}_k})(J_0)\right] \\ &= T_{\bar{\mu}_0}(J_{\bar{\pi}_1}) \leq T_{\bar{\mu}_0}(J^*) + \alpha\varepsilon_2 \leq T(J^*) + (\varepsilon_1 + \alpha\varepsilon_2). \end{aligned}$$

Since $J^* \leq J_{\bar{\pi}}$ and $\varepsilon_1$ and $\varepsilon_2$ can be taken arbitrarily small, it follows that

$$J^* \leq T(J^*).$$

Hence $J^* = T(J^*)$.

Assume that $J' \in F$ satisfies $J' \geq J_0$ and $J' \geq T(J')$. Let $\{\varepsilon_k\}$ be any sequence with $\varepsilon_k > 0$ and consider a policy $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots) \in \Pi$ such that

$$T_{\bar{\mu}_k}(J') \leq T(J') + \varepsilon_k, \qquad k = 0, 1, \ldots.$$

We have, from I.2,

$$\begin{aligned} J^* &= \inf_{\pi\in\Pi} \lim_{k\to\infty} (T_{\mu_0} \cdots T_{\mu_k})(J_0) \\ &\leq \inf_{\pi\in\Pi} \liminf_{k\to\infty} (T_{\mu_0} \cdots T_{\mu_k})(J') \\ &\leq \liminf_{k\to\infty} (T_{\bar{\mu}_0} \cdots T_{\bar{\mu}_k})(J') \\ &\leq \liminf_{k\to\infty} (T_{\bar{\mu}_0} \cdots T_{\bar{\mu}_{k-1}})[T(J') + \varepsilon_k] \\ &\leq \liminf_{k\to\infty} (T_{\bar{\mu}_0} \cdots T_{\bar{\mu}_{k-2}} T_{\bar{\mu}_{k-1}})(J' + \varepsilon_k) \\ &\leq \liminf_{k\to\infty} [(T_{\bar{\mu}_0} \cdots T_{\bar{\mu}_{k-1}})(J') + \alpha^k \varepsilon_k] \\ &\;\;\vdots \\ &\leq \lim_{k\to\infty}\left[T(J') + \left(\sum_{i=0}^{k} \alpha^i \varepsilon_i\right)\right] \leq J' + \left(\sum_{i=0}^{\infty} \alpha^i \varepsilon_i\right). \end{aligned}$$

Since we may choose $\sum_{i=0}^{\infty} \alpha^i \varepsilon_i$ as small as desired, it follows that $J^* \leq J'$. Q.E.D.

The following counterexamples show that I.1 and I.2 are essential in order for Proposition 5.2 to hold.

COUNTEREXAMPLE 1 Take $S=\{0,1\}$, $C=U(0)=U(1)=(-1,0]$, $J_0(0)=J_0(1)=-1$, $H(0,u,J)=u$ if $J(1)\leq -1$, $H(0,u,J)=0$ if $J(1)>-1$, and $H(1,u,J)=u$. Then $(T_{\mu_0}\cdots T_{\mu_{N-1}})(J_0)(0)=0$ and $(T_{\mu_0}\cdots T_{\mu_{N-1}})(J_0)(1)=\mu_0(1)$ for $N\geq 1$. Thus $J^*(0)=0$, $J^*(1)=-1$, while $T(J^*)(0)=-1$, $T(J^*)(1)=-1$, and hence $J^*\neq T(J^*)$. Notice also that $J_0$ is a fixed point of $T$, while $J_0\leq J^*$ and $J_0\neq J^*$, so the second part of Proposition 5.2 fails when $J_0=J'$. Here I and I.1 are satisfied, but I.2 is violated.

COUNTEREXAMPLE 2 Take $S=\{0,1\}$, $C=U(0)=U(1)=\{0\}$, $J_0(0)=J_0(1)=0$, $H(0,0,J)=0$ if $J(1)<\infty$, $H(0,0,J)=\infty$ if $J(1)=\infty$, $H(1,0,J)=J(1)+1$. Then $(T_{\mu_0}\cdots T_{\mu_{N-1}})(J_0)(0)=0$ and $(T_{\mu_0}\cdots T_{\mu_{N-1}})(J_0)(1)=N$. Thus $J^*(0)=0$, $J^*(1)=\infty$. On the other hand, we have $T(J^*)(0)=T(J^*)(1)=\infty$ and $J^*\neq T(J^*)$. Here I and I.2 are satisfied, but I.1 is violated.

As a corollary to Proposition 5.1 we obtain the following.

**Corollary 5.2.1** Let I, I.1, and I.2 hold. Then for every stationary policy $\pi=(\mu,\mu,\ldots)$, we have

$$J_\mu = T_\mu(J_\mu).$$

Furthermore, if $J'\in F$ is such that $J'\geq J_0$ and $J'\geq T_\mu(J')$, then $J'\geq J_\mu$.

*Proof* Consider the variation of our problem where the control constraint set is $U_\mu(x)=\{\mu(x)\}$ rather than $U(x)$ for $\forall x\in S$. Application of Proposition 5.2 yields the result. Q.E.D.

We now provide the counterpart of Proposition 5.2 under Assumption D.

**Proposition 5.3** Let D and D.1 hold. Then

$$J^* = T(J^*).$$

Furthermore, if $J'\in F$ is such that $J'\leq J_0$ and $J'\leq T(J')$, then $J'\leq J^*$.

*Proof* We first show the following lemma.

**Lemma 5.1** Let D hold. Then

$$J^* = \lim_{N\to\infty} J_N^*, \tag{15}$$

where $J_N^*$ is the optimal cost function for the $N$-stage problem.

*Proof* Clearly we have $J^*\leq J_N^*$ for all $N$, and hence $J^*\leq \lim_{N\to\infty} J_N^*$. Also, for all $\pi=(\mu_0,\mu_1,\ldots)\in\Pi$, we have

$$(T_{\mu_0}\cdots T_{\mu_{N-1}})(J_0)\geq J_N^*,$$

and by taking the limit on both sides we obtain $J_\pi\geq \lim_{N\to\infty} J_N^*$, and hence $J^*\geq \lim_{N\to\infty} J_N^*$. Q.E.D.

*Proof* (continued) We return now to the proof of Proposition 5.3. An argument entirely similar to the one of the proof of Lemma 5.1 shows that under D we have for all $x \in S$

$$\lim_{N\to\infty} \inf_{u\in U(x)} H(x,u,J_N^*) = \inf_{u\in U(x)} \lim_{N\to\infty} H(x,u,J_N^*). \tag{16}$$

Using D.1, this equation yields

$$\lim_{N\to\infty} T(J_N^*) = T\left(\lim_{N\to\infty} J_N^*\right). \tag{17}$$

Since D and D.1 are equivalent to Assumption F.1′ of Chapter 3, by Corollary 3.1.1 we have $J_N^* = T^N(J_0)$, from which we conclude that $T(J_N^*) = T^{N+1}(J_0) = J_{N+1}^*$. Combining this relation with (15) and (17), we obtain $J^* = T(J^*)$.

To complete the proof, let $J' \in F$ be such that $J' \le J_0$ and $J' \le T(J')$. Then we have

$$\begin{aligned} J^* &= \inf_{\pi\in\Pi} \lim_{N\to\infty} (T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0) \\ &\ge \lim_{N\to\infty} \inf_{\pi\in\Pi} (T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0) \\ &\ge \lim_{N\to\infty} \inf_{\pi\in\Pi} (T_{\mu_0} \cdots T_{\mu_{N-1}})(J') \ge \lim_{N\to\infty} T^N(J') \ge J'. \end{aligned}$$

Hence $J^* \ge J'$. Q.E.D.

In Counterexamples 1 and 2 of Section 3.2, D is satisfied but D.1 is violated. In both cases we have $J^* \ne T(J^*)$, as the reader can easily verify.

A cursory examination of the proof of Proposition 5.3 reveals that the only point where we used D.1 was in establishing the relations

$$\lim_{N\to\infty} T(J_N^*) = T(\lim_{N\to\infty} J_N^*)$$

and $J_N^* = T^N(J_0)$. If these relations can be established independently, then the result of Proposition 5.3 follows. In this manner we obtain the following corollary.

**Corollary 5.3.1** Let D hold and assume that D.2 holds, $S$ is a finite set, and $J^*(x) > -\infty$ for all $x \in S$. Then $J^* = T(J^*)$. Furthermore, if $J' \in F$ is such that $J' \le J_0$ and $J' \le T(J')$, then $J' \le J^*$.

*Proof* A nearly verbatim repetition of the proof of Proposition 3.1(b) shows that, under D, D.2, and the assumption that $J^*(x) > -\infty$ for all $x \in S$, we have $J_N^* = T^N(J_0)$ for all $N = 1, 2, \ldots$. We will show that

$$\lim_{N\to\infty} H(x,u,J_N^*) = H\left(x,u,\lim_{N\to\infty} J_N^*\right) \qquad \forall x \in S, \quad u \in U(x).$$

Then using (16) we obtain (17), and the result follows as in the proof of Proposition 5.3. Assume the contrary, i.e., that for some $\tilde{x} \in S$, $\tilde{u} \in U(\tilde{x})$, and

$\varepsilon > 0$, there holds

$$H(\tilde{x}, \tilde{u}, J_k^*) - \varepsilon > H\left(\tilde{x}, \tilde{u}, \lim_{N\to\infty} J_N^*\right), \qquad k = 1, 2, \ldots .$$

From the finiteness of $S$ and the fact that $J^*(x) = \lim_{N\to\infty} J_N^*(x) > -\infty$ for all $x$, we know that for some positive integer $\bar{k}$

$$J_k^* - (\varepsilon/\alpha) \leq \lim_{N\to\infty} J_N^* \qquad \forall k \geq \bar{k}.$$

By using D.2 we obtain for all $k \geq \bar{k}$

$$H(\tilde{x}, \tilde{u}, J_k^*) - \varepsilon \leq H(\tilde{x}, \tilde{u}, J_k^* - (\varepsilon/\alpha)) \leq H\left(\tilde{x}, \tilde{u}, \lim_{N\to\infty} J_N^*\right),$$

which contradicts the earlier inequality. Q.E.D.

Similarly, as under Assumption I, we have the following corollary.

**Corollary 5.3.2** Let D and D.1 hold. Then for every stationary policy $\pi = (\mu, \mu, \ldots)$, we have

$$J_\mu = T_\mu(J_\mu).$$

Furthermore, if $J' \in F$ is such that $J' \leq J_0$ and $J' \leq T_\mu(J')$, then $J' \leq J_\mu$.

It is worth noting that Propositions 5.2 and 5.3 can form the basis for computation of $J^*$ when the state space $S$ is a finite set with $n$ elements denoted by $x_1, x_2, \ldots, x_n$. It follows from Proposition 5.2 that, under I, I.1, and I.2, $J^*(x_1), \ldots, J^*(x_n)$ solve the problem

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{n} \lambda_i \\ &\text{subject to} \quad \lambda_i \geq \inf_{u \in U(x_i)} H(x_i, u, J_\lambda), \qquad i = 1, \ldots, n, \\ &\qquad\qquad\quad \lambda_i \geq J_0(x_i), \qquad i = 1, \ldots, n, \end{aligned}$$

where $J_\lambda$ is the function taking values $J_\lambda(x_i) = \lambda_i$, $i = 1, \ldots, n$. Under D and D.1, or D, D.2, and the assumption that $J^*(x) > -\infty$ for $\forall x \in S$, the corresponding problem is

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{n} \lambda_i \\ &\text{subject to} \quad \lambda_i \leq H(x_i, u, J_\lambda), \qquad i = 1, \ldots, n, \quad u \in U(x_i) \\ &\qquad\qquad\quad \lambda_i \leq J_0(x_i), \qquad i = 1, \ldots, n. \end{aligned}$$

When $U(x_i)$ is also a finite set for all $i$, then the preceding problem becomes a finite-dimensional (possibly nonlinear) programming problem.

## 5.3 Characterization of Optimal Policies

We have the following necessary and sufficient conditions for optimality of a stationary policy.

**Proposition 5.4** Let I, I.1, and I.2 hold. Then a stationary policy $\pi^* = (\mu^*, \mu^*, \ldots)$ is optimal if and only if

$$T_{\mu^*}(J^*) = T(J^*). \tag{18}$$

Furthermore, if for each $x \in S$, there exists a policy which is optimal at $x$, then there exists a stationary optimal policy.

*Proof* If $\pi^*$ is optimal, then $J_{\mu^*} = J^*$ and the result follows from Proposition 5.2 and Corollary 5.2.1. Conversely, if $T_{\mu^*}(J^*) = T(J^*)$, then since $J^* = T(J^*)$ (Proposition 5.2), it follows that $T_{\mu^*}(J^*) = J^*$. Hence by Corollary 5.2.1, $J_{\mu^*} \leq J^*$ and it follows that $\pi^*$ is optimal.

If $\pi_x^* = (\mu_{0,x}^*, \mu_{1,x}^*, \ldots)$ is optimal at $x \in S$, we have, from I.1,

$$\begin{aligned} J^*(x) = J_{\pi_x^*}(x) &= \lim_{k\to\infty} (T_{\mu_{0,x}^*} \cdots T_{\mu_{k,x}^*})(J_0)(x) \\ &= T_{\mu_{0,x}^*}\left[\lim_{k\to\infty} (T_{\mu_{1,x}^*} \cdots T_{\mu_{k,x}^*})(J_0)\right](x) \\ &\geq T_{\mu_{0,x}^*}(J^*)(x) \geq T(J^*)(x) = J^*(x). \end{aligned}$$

Hence $T_{\mu_{0,x}^*}(J^*)(x) = T(J^*)(x)$ for all $x \in S$. Define $\mu^* \in M$ by $\mu^*(x) = \mu_{0,x}^*(x)$. Then $T_{\mu^*}(J^*) = T(J^*)$ and, by the result just proved, the stationary policy $(\mu^*, \mu^*, \ldots)$ is optimal. Q.E.D.

**Proposition 5.5** Let D and D.1 hold. Then a stationary policy $\pi^* = (\mu^*, \mu^*, \ldots)$ is optimal if and only if

$$T_{\mu^*}(J_{\mu^*}) = T(J_{\mu^*}). \tag{19}$$

*Proof* If $\pi^*$ is optimal, then $J_{\mu^*} = J^*$, and the result follows from Proposition 5.3 and Corollary 5.3.2. Conversely, if $T_{\mu^*}(J_{\mu^*}) = T(J_{\mu^*})$, then we obtain, from Corollary 5.3.2, that $J_{\mu^*} = T(J_{\mu^*})$, and Proposition 5.3 yields $J_{\mu^*} \leq J^*$. Hence $\pi^*$ is optimal. Q.E.D.

Examples where $\pi^*$ satisfies (18) or (19) but is not optimal under D or I, respectively, are given in DPSC, Section 6.4.

Proposition 5.4 shows that there exists a stationary optimal policy if and only if the infimum in the optimality equation

$$J^*(x) = \inf_{u \in U(x)} H(x, u, J^*)$$

is attained for every $x \in S$. When the infimum is not attained for some $x \in S$, the optimality equation can still be used to yield an $\varepsilon$-optimal policy, which can be taken to be stationary whenever the scalar $\alpha$ in I.2 is strictly less than one. This is shown in the following proposition.

**Proposition 5.6** Let I, I.1, and I.2 hold. Then:

(a) If $\varepsilon > 0$, $\{\varepsilon_i\}$ satisfies $\sum_{k=0}^{\infty} \alpha^k \varepsilon_k = \varepsilon$, $\varepsilon_i > 0$, $i = 0, 1, \ldots$, and $\pi^* = (\mu_0^*, \mu_1^*, \ldots) \in \Pi$ is such that

$$T_{\mu_k^*}(J^*) \leq T(J^*) + \varepsilon_k, \qquad k = 0, 1, \ldots,$$

then

$$J^* \leq J_{\pi^*} \leq J^* + \varepsilon.$$

(b) If $\varepsilon > 0$, the scalar $\alpha$ in I.2 is strictly less than one, and $\mu^* \in M$ is such that

$$T_{\mu^*}(J^*) \leq T(J^*) + \varepsilon(1 - \alpha),$$

then

$$J^* \leq J_{\mu^*} \leq J^* + \varepsilon.$$

*Proof* (a) Since $T(J^*) = J^*$, we have $T_{\mu_k^*}(J^*) \leq J^* + \varepsilon_k$, and applying $T_{\mu_{k-1}^*}$ to both sides we obtain

$$(T_{\mu_{k-1}^*} T_{\mu_k^*})(J^*) \leq T_{\mu_{k-1}^*}(J^*) + \alpha\varepsilon_k \leq J^* + (\varepsilon_{k-1} + \alpha\varepsilon_k).$$

Applying $T_{\mu_{k-2}^*}$ throughout and repeating the process, we obtain, for every $k = 1, 2, \ldots,$

$$(T_{\mu_0^*} \cdots T_{\mu_k^*})(J^*) \leq J^* + \left(\sum_{i=0}^{k} \alpha^i \varepsilon_i\right).$$

Since $J_0 \leq J^*$, it follows that

$$(T_{\mu_0^*} \cdots T_{\mu_k^*})(J_0) \leq J^* + \left(\sum_{i=0}^{k} \alpha^i \varepsilon_i\right), \qquad k = 1, 2, \ldots.$$

By taking the limit as $k \to \infty$, we obtain $J_{\pi^*} \leq J^* + \varepsilon$.

(b) This part is proved by taking $\varepsilon_k = \varepsilon(1 - \alpha)$ and $\mu_k^* = \mu^*$ for all $k$ in the preceding proof. Q.E.D.

A weak counterpart of part (a) of Proposition 5.6 under D is given in Corollary 5.7.1. We have been unable to obtain a counterpart of part (b) or conditions for existence of a stationary optimal policy under D.

## 5.4 Convergence of the Dynamic Programming Algorithm—Existence of Stationary Optimal Policies

The DP algorithm consists of successive generation of the functions $T(J_0), T^2(J_0), \ldots$. Under Assumption I we have $T^k(J_0) \leq T^{k+1}(J_0)$ for all $k$, while under Assumption D we have $T^{k+1}(J_0) \leq T^k(J_0)$ for all $k$. Thus we can define a function $J_\infty \in F$ by

$$J_\infty(x) = \lim_{N\to\infty} T^N(J_0)(x) \qquad \forall x \in S. \tag{20}$$

We would like to investigate the question whether $J_\infty = J^*$. When Assumption D holds, the following proposition shows that we have $J_\infty = J^*$ under mild assumptions.

**Proposition 5.7** Let D hold and assume that either D.1 holds or else $J_N^* = T^N(J_0)$ for all $N$, where $J_N^*$ is the optimal cost function of the $N$-stage problem. Then

$$J_\infty = J^*.$$

*Proof* By Lemma 5.1 we have that D implies $J^* = \lim_{N\to\infty} J_N^*$. Corollary 3.1.1 shows also that under our assumptions $J_N^* = T^N(J_0)$. Hence $J^* = \lim_{N\to\infty} T^N(J_0) = J_\infty$. Q.E.D.

The following corollary is a weak counterpart of Proposition 5.1 and part (a) of Proposition 5.6 under D.

**Corollary 5.7.1** Let D hold and assume that D.2 holds, $S$ is a finite set, and $J^*(x) > -\infty$ for all $x \in S$. Then for any $\varepsilon > 0$, there exists an $\varepsilon$-optimal policy, i.e., a $\pi_\varepsilon \in \Pi$ such that

$$J^* \leq J_{\pi_\varepsilon} \leq J^* + \varepsilon$$

*Proof* For each $N$, denote $\varepsilon_N = \varepsilon/2(1 + \alpha + \cdots + \alpha^{N-1})$ and let $\pi_N = \{\mu_0^N, \mu_1^N, \ldots, \mu_{N-1}^N, \mu, \mu, \ldots\}$ be such that $\mu \in M$, and for $k = 0, 1, \ldots, N-1$, $\mu_k^N \in M$ and

$$(T_{\mu_k^N} T^{N-k-1})(J_0) \leq T^{N-k}(J_0) + \varepsilon_N.$$

We have $T_{\mu_{N-1}^N}(J_0) \leq T(J_0) + \varepsilon_N$, and applying $T_{\mu_{N-2}^N}$ to both sides, we obtain

$$(T_{\mu_{N-2}^N} T_{\mu_{N-1}^N})(J_0) \leq (T_{\mu_{N-2}^N} T)(J_0) + \alpha\varepsilon_N \leq T^2(J_0) + (1 + \alpha)\varepsilon_N.$$

Continuing in the same manner, we have

$$(T_{\mu_0^N} \cdots T_{\mu_{N-1}^N})(J_0) \leq T^N(J_0) + (1 + \alpha + \cdots + \alpha^{N-1})\varepsilon_N,$$

from which we obtain, for $N = 0, 1, \ldots,$

$$J_{\pi_N} \leq T^N(J_0) + (\varepsilon/2).$$

As in the proof of Corollary 5.3.1 our assumptions imply that $J_N^* = T^N(J_0)$ for all $N$, so by Proposition 5.7, $\lim_{N\to\infty} T^N(J_0) = J^*$. Let $\bar{N}$ be such that $T^{\bar{N}}(J_0) \leq J^* + (\varepsilon/2)$. Such an $\bar{N}$ exists by the finiteness of $S$ and the fact that $J^*(x) > -\infty$ for all $x \in S$. Then we obtain $J_{\pi_{\bar{N}}} \leq J^* + \varepsilon$, and $\pi_{\bar{N}}$ is the desired policy. Q.E.D.

Under Assumptions I, I.1, and I.2, the equality $J_\infty = J^*$ may fail to hold even in simple deterministic optimal control problems, as shown in the following counterexample.

COUNTEREXAMPLE 3 Let $S = [0, \infty)$, $C = U(x) = (0, \infty)$ for $\forall x \in S$, $J_0(x) = 0$ for $\forall x \in S$, and

$$H(x, u, J) = x + J(2x + u) \qquad \forall x \in S, \quad u \in U(x).$$

Then it is easy to verify that

$$J^*(x) = \inf_{\pi \in \Pi} J_\pi(x) = \infty \qquad \forall x \in S,$$

while

$$T^N(J_0)(0) = 0, \qquad N = 1, 2, \ldots.$$

Hence $J_\infty(0) = 0$ and $J_\infty(0) < J^*(0)$.

In this example, we have $J^*(x) = \infty$ for all $x \in S$. Other examples exist where $J^* \neq J_\infty$ and $J^*(x) < \infty$ for all $x \in S$ (see [S14, p. 880]). The following preliminary result shows that in order to have $J_\infty = J^*$, it is necessary and sufficient to have $J_\infty = T(J_\infty)$.

**Proposition 5.8** Let I, I.1, and I.2 hold. Then

$$J_\infty \leq T(J_\infty) \leq T(J^*) = J^*. \tag{21}$$

Furthermore, the equalities

$$J_\infty = T(J_\infty) = T(J^*) = J^* \tag{22}$$

hold if and only if

$$J_\infty = T(J_\infty). \tag{23}$$

*Proof* Clearly we have $J_\infty \leq J_\pi$ for all $\pi \in \Pi$, and it follows that $J_\infty \leq J^*$. Furthermore, by Proposition 5.2 we have $T(J^*) = J^*$. Also, we have, for all $k \geq 0$,

$$T(J_\infty) = \inf_{u \in U(x)} H(x, u, J_\infty) \geq \inf_{u \in U(x)} H[x, u, T^k(J_0)] = T^{k+1}(J_0).$$

Taking the limit of the right side, we obtain $T(J_\infty) \geq J_\infty$, which, combined with $J_\infty \leq J^*$ and $T(J^*) = J^*$, proves (21). If (22) holds, then (23) also holds.

Conversely, let (23) hold. Then since we have $J_\infty \geq J_0$, we see from Proposition 5.2 that $J_\infty \geq J^*$, which combined with (21) proves (22). Q.E.D.

In what follows we provide a necessary and sufficient condition for $J_\infty = T(J_\infty)$ [and hence also (22)] to hold under Assumptions I, I.1, and I.2. We subsequently obtain a useful sufficient condition for $J_\infty = T(J_\infty)$ to hold, which at the same time guarantees the existence of a stationary optimal policy.

For any $J \in F$, we denote by $E(J)$ the *epigraph* of $J$, i.e., the subset of $SR$ given by

$$E(J) = \{(x, \lambda) | J(x) \leq \lambda\}. \tag{24}$$

Under I we have $T^k(J_0) \leq T^{k+1}(J_0)$ for all $k$ and also $J_\infty = \lim_{k\to\infty} T^k(J_0)$, so it follows easily that

$$E(J_\infty) = \bigcap_{k=1}^{\infty} E[T^k(J_0)]. \tag{25}$$

Consider for each $k \geq 1$ the subset $C_k$ of $SCR$ given by

$$C_k = \{(x, u, \lambda) | H[x, u, T^{k-1}(J_0)] \leq \lambda,\ x \in S,\ u \in U(x)\}. \tag{26}$$

Denote by $P(C_k)$ the projection of $C_k$ on $SR$, i.e.,

$$P(C_k) = \{(x, \lambda) | \exists u \in U(x) \text{ s.t. } (x, u, \lambda) \in C_k\}.^\dagger \tag{27}$$

Consider also the set

$$\overline{P(C_k)} = \{(x, \lambda) | \exists \{\lambda_n\} \text{ s.t. } \lambda_n \to \lambda,\ (x, \lambda_n) \in P(C_k),\ n = 0, 1, \ldots\}. \tag{28}$$

The set $\overline{P(C_k)}$ is obtained from $P(C_k)$ by adding for each $x$ the point $[x, \bar{\lambda}(x)]$ where $\bar{\lambda}(x)$ is the perhaps missing end point of the half-line $\{\lambda | (x, \lambda) \in P(C_k)\}$. We have the following lemma.

**Lemma 5.2** Let I hold. Then for all $k \geq 1$

$$P(C_k) \subset \overline{P(C_k)} = E[T^k(J_0)]. \tag{29}$$

Furthermore, we have

$$P(C_k) = \overline{P(C_k)} = E[T^k(J_0)] \tag{30}$$

if and only if the infimum is attained for each $x \in S$ in the equation

$$T^k(J_0)(x) = \inf_{u \in U(x)} H[x, u, T^{k-1}(J_0)]. \tag{31}$$

† The symbol ∃ means "there exists" and the initials "s.t." stand for "such that."

*Proof* If $(x, \lambda) \in E[T^k(J_0)]$, we have

$$T^k(J_0)(x) = \inf_{u \in U(x)} H[x, u, T^{k-1}(J_0)] \leq \lambda.$$

Let $\{\varepsilon_n\}$ be a sequence such that $\varepsilon_n > 0$, $\varepsilon_n \to 0$, and let $\{u_n\} \subset U(x)$ be a sequence such that

$$H[x, u_n, T^{k-1}(J_0)] \leq T^k(J_0)(x) + \varepsilon_n \leq \lambda + \varepsilon_n.$$

Then $(x, u_n, \lambda + \varepsilon_n) \in C_k$ and $(x, \lambda + \varepsilon_n) \in P(C_k)$ for all $n$. Since $\lambda + \varepsilon_n \to \lambda$, by (28) we obtain $(x, \lambda) \in \overline{P(C_k)}$. Hence

$$E[T^k(J_0)] \subset \overline{P(C_k)}. \tag{32}$$

Conversely, let $(x, \lambda) \in \overline{P(C_k)}$. Then by (26)–(28) there exists a sequence $\{\lambda_n\}$ with $\lambda_n \to \lambda$ and a corresponding sequence $\{u_n\} \subset U(x)$ such that

$$T^k(J_0)(x) \leq H[x, u_n, T^{k-1}(J_0)] \leq \lambda_n.$$

Taking the limit as $n \to \infty$, we obtain $T^k(J_0)(x) \leq \lambda$ and $(x, \lambda) \in E[T^k(J_0)]$. Hence

$$\overline{P(C_k)} \subset E[T^k(J_0)],$$

and using (32) we obtain (29).

To prove that (30) is equivalent to the attainment of the infimum in (31), assume first that the infimum is attained by $\mu_{k-1}^*(x)$ for each $x \in S$. Then for each $(x, \lambda) \in E[T^k(J_0)]$,

$$H[x, \mu_{k-1}^*(x), T^{k-1}(J_0)] \leq \lambda,$$

which implies by (27) that $(x, \lambda) \in P(C_k)$. Hence $E[T^k(J_0)] \subset P(C_k)$ and, in view of (29), we obtain (30). Conversely, if (30) holds, we have $[x, T^k(J_0)(x)] \in P(C_k)$ for every $x$ for which $T^k(J_0)(x) < \infty$. Then by (26) and (27), there exists a $\mu_{k-1}^*(x) \in U(x)$ such that

$$H[x, \mu_{k-1}^*(x), T^{k-1}(J_0)] \leq T^k(J_0)(x) = \inf_{u \in U(x)} H[x, u, T^{k-1}(J_0)].$$

Hence the infimum in (31) is attained for all $x$ for which $T^k(J_0)(x) < \infty$. It is also trivially attained by all $u \in U(x)$ whenever $T^k(J_0)(x) = \infty$, and the proof is complete. Q.E.D.

Consider now the set $\bigcap_{k=1}^{\infty} C_k$ and define similarly as in (27) and (28) the sets

$$P\left(\bigcap_{k=1}^{\infty} C_k\right) = \left\{(x, \lambda) \mid \exists u \in U(x) \text{ s.t. } (x, u, \lambda) \in \bigcap_{k=1}^{\infty} C_k\right\}, \tag{33}$$

$$\overline{P\left(\bigcap_{k=1}^{\infty} C_k\right)} = \left\{(x, \lambda) \mid \exists \{\lambda_n\} \text{ s.t. } \lambda_n \to \lambda,\ (x, \lambda_n) \in P\left(\bigcap_{k=1}^{\infty} C_k\right)\right\}. \tag{34}$$

Using (25) and Lemma 5.2, it is easy to see that

$$P\left(\bigcap_{k=1}^{\infty} C_k\right) \subset \bigcap_{k=1}^{\infty} P(C_k) \subset \bigcap_{k=1}^{\infty} \overline{P(C_k)} = \bigcap_{k=1}^{\infty} E[T^k(J_0)] = E(J_\infty), \tag{35}$$

$$\overline{P\left(\bigcap_{k=1}^{\infty} C_k\right)} \subset \bigcap_{k=1}^{\infty} \overline{P(C_k)} = \bigcap_{k=1}^{\infty} E[T^k(J_0)] = E(J_\infty). \tag{36}$$

We have the following proposition.

**Proposition 5.9** Let I, I.1, and I.2 hold. Then:

(a) We have $J_\infty = T(J_\infty)$ (equivalently $J_\infty = J^*$) if and only if

$$\overline{P\left(\bigcap_{k=1}^{\infty} C_k\right)} = \bigcap_{k=1}^{\infty} \overline{P(C_k)}. \tag{37}$$

(b) We have $J_\infty = T(J_\infty)$ (equivalently $J_\infty = J^*$), and the infimum in

$$J_\infty(x) = \inf_{u \in U(x)} H(x, u, J_\infty) \tag{38}$$

is attained for each $x \in S$ (equivalently there exists a stationary optimal policy) if and only if

$$P\left(\bigcap_{k=1}^{\infty} C_k\right) = \bigcap_{k=1}^{\infty} \overline{P(C_k)}. \tag{39}$$

*Proof* (a) Assume $J_\infty = T(J_\infty)$ and let $(x, \lambda)$ be in $E(J_\infty)$, i.e.,

$$\inf_{u \in U(x)} H(x, u, J_\infty) = J_\infty(x) \le \lambda.$$

Let $\{\varepsilon_n\}$ be any sequence with $\varepsilon_n > 0$, $\varepsilon_n \to 0$. Then there exists a sequence $\{u_n\}$ such that

$$H(x, u_n, J_\infty) \le \lambda + \varepsilon_n, \qquad n = 1, 2, \ldots,$$

and so

$$H[x, u_n, T^{k-1}(J_0)] \le \lambda + \varepsilon_n, \qquad k, n = 1, 2, \ldots.$$

It follows that $(x, u_n, \lambda + \varepsilon_n) \in C_k$ for all $k$, $n$ and $(x, u_n, \lambda + \varepsilon_n) \in \bigcap_{k=1}^{\infty} C_k$ for all $n$. Hence $(x, \lambda + \varepsilon_n) \in P(\bigcap_{k=1}^{\infty} C_k)$ for all $n$, and since $\lambda + \varepsilon_n \to \lambda$, we obtain $(x, \lambda) \in \overline{P(\bigcap_{k=1}^{\infty} C_k)}$. Therefore

$$E(J_\infty) \subset \overline{P\left(\bigcap_{k=1}^{\infty} C_k\right)},$$

and using (36) we obtain (37).

Conversely, let (37) hold. Then by (36) we have $\overline{P(\bigcap_{k=1}^{\infty} C_k)} = E(J_\infty)$. Let $x \in S$ be such that $J_\infty(x) < \infty$. Then $[x, J_\infty(x)] \in \overline{P(\bigcap_{k=1}^{\infty} C_k)}$, and there exists a sequence $\{\lambda_n\}$ with $\lambda_n \to J_\infty(x)$ and a sequence $\{u_n\} \subset U(x)$ such that

$$H[x, u_n, T^{k-1}(J_0)] \leq \lambda_n, \qquad k, n = 1, 2, \ldots.$$

Taking the limit with respect to $k$ and using I.1, we obtain

$$H(x, u_n, J_\infty) \leq \lambda_n,$$

and since $T(J_\infty)(x) \leq H(x, u_n, J_\infty)$, it follows that

$$T(J_\infty)(x) \leq \lambda_n.$$

Taking the limit as $n \to \infty$, we obtain

$$T(J_\infty)(x) \leq J_\infty(x)$$

for all $x \in S$ such that $J_\infty(x) < \infty$. Since the preceding inequality holds also for all $x \in S$ with $J_\infty(x) = \infty$, we have

$$T(J_\infty) \leq J_\infty.$$

On the other hand, by Proposition 5.8, we have

$$J_\infty \leq T(J_\infty).$$

Combining the two inequalities, we obtain $J_\infty = T(J_\infty)$.

(b) Assume $J_\infty = T(J_\infty)$ and that the infimum in (38) is attained for each $x \in S$. Then there exists a function $\mu^* \in M$ such that for each $(x, \lambda) \in E(J_\infty)$

$$H[x, \mu^*(x), J_\infty] \leq \lambda.$$

Hence $H[x, \mu^*(x), T^{k-1}(J_0)] \leq \lambda$ for $k = 1, 2, \ldots$, and we have $[x, \mu^*(x), \lambda] \in \bigcap_{k=1}^\infty C_k$. As a result, $(x, \lambda) \in P(\bigcap_{k=1}^\infty C_k)$. Hence

$$E(J_\infty) \subset P\left(\bigcap_{k=1}^\infty C_k\right),$$

and (39) follows from (35).

Conversely, let (39) hold. We have for all $x \in S$ with $J_\infty(x) < \infty$ that

$$[x, J_\infty(x)] \in E(J_\infty) = P\left(\bigcap_{k=1}^\infty C_k\right).$$

Thus there exists a $\mu^*(x) \in U(x)$ such that

$$[x, \mu^*(x), J_\infty(x)] \in \bigcap_{k=1}^\infty C_k,$$

from which we conclude that

$$H[x, \mu^*(x), T^{k-1}(J_0)] \leq J_\infty(x), \quad k = 0, 1, \ldots.$$

Taking the limit and using I.1, we see that

$$T(J_\infty)(x) \le H[x, \mu^*(x), J_\infty] \le J_\infty(x).$$

It follows that $T(J_\infty) \le J_\infty$, and since $J_\infty \le T(J_\infty)$ by Proposition 5.8, we finally obtain $J_\infty = T(J_\infty)$. Furthermore, the last inequality shows that $\mu^*(x)$ attains the infimum in (38) when $J_\infty(x) < \infty$. When $J_\infty(x) = \infty$, every $u \in U(x)$ attains the infimum, and the proof is complete. Q.E.D.

In view of Proposition 5.8, the equality $J_\infty = T(J_\infty)$ is equivalent to the validity of interchanging infimum and limit as follows

$$J_\infty = \lim_{k\to\infty} \inf_{\pi\in\Pi} (T_{\mu_0} \cdots T_{\mu_k})(J_0) = \inf_{\pi\in\Pi} \lim_{k\to\infty} (T_{\mu_0} \cdots T_{\mu_k})(J_0) = J^*.$$

Thus Proposition 5.9 states that interchanging infimum and limit is in fact equivalent to the validity of interchanging projection and intersection in the manner of (37) or (39).

The following proposition provides a compactness assumption which guarantees that (39) holds.

**Proposition 5.10** Let I, I.1, and I.2 hold and let the control space $C$ be a Hausdorff space. Assume that there exists a nonnegative integer $\bar{k}$ such that for each $x \in S$, $\lambda \in R$, and $k \ge \bar{k}$, the set

$$U_k(x, \lambda) = \{u \in U(x) | H[x, u, T^k(J_0)] \le \lambda\} \tag{40}$$

is compact. Then

$$P\left(\bigcap_{k=1}^{\infty} C_k\right) = \bigcap_{k=1}^{\infty} \overline{P(C_k)} \tag{41}$$

and (by Propositions 5.8 and 5.9)

$$J_\infty = T(J_\infty) = T(J^*) = J^*.$$

Furthermore, there exists a stationary optimal policy.

*Proof* By (35) it will be sufficient to show that

$$P\left(\bigcap_{k=1}^{\infty} C_k\right) \supset \bigcap_{k=1}^{\infty} P(C_k), \qquad \bigcap_{k=1}^{\infty} P(C_k) = \bigcap_{k=1}^{\infty} \overline{P(C_k)}. \tag{42}$$

Let $(x, \lambda)$ be in $\bigcap_{k=1}^{\infty} P(C_k)$. Then there exists a sequence $\{u_n\} \subset U(x)$ such that

$$H[x, u_n, T^k(J_0)] \le H[x, u_n, T^n(J_0)] \le \lambda \qquad \forall n \ge k,$$

or equivalently

$$u_n \in U_k(x, \lambda) \qquad \forall n \ge k, \quad k = 0, 1, \ldots.$$

Since $U_k(x, \lambda)$ is compact for $k \geq \bar{k}$, it follows that the sequence $\{u_n\}$ has an accumulation point $\bar{u}$ and

$$\bar{u} \in U_k(x, \lambda) \qquad \forall k \geq \bar{k}.$$

But $U_0(x, \lambda) \supset U_1(x, \lambda) \supset \ldots$, so $\bar{u} \in U_k(x, \lambda)$ for $k = 0, 1, \ldots$. Hence

$$H[x, \bar{u}, T^k(J_0)] \leq \lambda, \qquad k = 0, 1, \ldots,$$

and $(x, \bar{u}, \lambda) \in \bigcap_{k=1}^{\infty} C_k$. It follows that $(x, \lambda) \in P(\bigcap_{k=1}^{\infty} C_k)$ and

$$P\left(\bigcap_{k=1}^{\infty} C_k\right) \supset \bigcap_{k=1}^{\infty} P(C_k).$$

Also, from the compactness of $U_k(x, \lambda)$ and the result of Lemma 3.1, it follows that the infimum in (31) is attained for every $x \in S$ and $k > \bar{k}$. By Lemma 5.2, $P(C_k) = \overline{P(C_k)}$ for $k > \bar{k}$, and since $P(C_1) \supset P(C_2) \supset \cdots$ and $\overline{P(C_1)} \supset \overline{P(C_2)} \supset \cdots$, we have

$$\bigcap_{k=1}^{\infty} P(C_k) = \bigcap_{k=1}^{\infty} \overline{P(C_k)}.$$

Thus (42) is proved. Q.E.D.

The following proposition shows also that a stationary optimal policy may be obtained in the limit by means of the DP algorithm.

**Proposition 5.11** Let the assumptions of Proposition 5.10 hold. Then:

(a) There exists a policy $\pi^* = (\mu_0^*, \mu_1^*, \ldots) \in \Pi$ attaining the infimum in the DP algorithm for all $k \geq \bar{k}$, i.e.,

$$(T_{\mu_k^*} T^k)(J_0) = T^{k+1}(J_0) \qquad \forall k \geq \bar{k}. \tag{43}$$

(b) For every policy $\pi^*$ satisfying (43), the sequence $\{\mu_k^*(x)\}$ has at least one accumulation point for each $x \in S$ with $J^*(x) < \infty$.

(c) If $\mu^*: S \to C$ is such that $\mu^*(x)$ is an accumulation point of $\{\mu_k^*(x)\}$ for all $x \in S$ with $J^*(x) < \infty$ and $\mu^*(x) \in U(x)$ for all $x \in S$ with $J^*(x) = \infty$, then the stationary policy $(\mu^*, \mu^*, \ldots)$ is optimal.

*Proof* (a) This follows from Lemma 3.1.

(b) For any $\pi^* = (\mu_0^*, \mu_1^*, \ldots)$ satisfying (43) and $x \in S$ such that $J^*(x) < \infty$, we have

$$H[x, \mu_n^*(x), T^k(J_0)] \leq H[x, \mu_n^*(x), T^n(J_0)] \leq J^*(x) \qquad \forall k \geq \bar{k}, \quad n \geq k.$$

Hence,

$$\mu_n^*(x) \in U_k[x, J^*(x)] \qquad \forall k \geq \bar{k}, \quad n \geq k.$$

Since $U_k[x, J^*(x)]$ is compact, $\{\mu_n^*(x)\}$ has at least one accumulation point. Furthermore, each accumulation point $\mu^*(x)$ of $\{\mu_n^*(x)\}$ belongs to $U(x)$ and satisfies

$$H[x, \mu^*(x), T^k(J_0)] \le J^*(x) \qquad \forall k \ge \bar{k}. \tag{44}$$

By taking the limit in (44) and using I.1, we obtain

$$H[x, \mu^*(x), J_\infty] = H[x, \mu^*(x), J^*] \le J^*(x)$$

for all $x \in S$ with $J^*(x) < \infty$. This relation holds trivially for all $x \in S$ with $J^*(x) = \infty$. Hence $T_{\mu^*}(J^*) \le J^* = T(J^*)$, which implies that $T_{\mu^*}(J^*) = T(J^*)$. It follows from Proposition 5.4 that $(\mu^*, \mu^*, \ldots)$ is optimal. Q.E.D.

The compactness of the sets $U_k(x, \lambda)$ of (40) may be verified in a number of special cases, some examples of which are given at the end of Section 3.2. Another example is the lower semicontinuous model of Section 8.3, whose infinite horizon version is treated in Corollary 9.17.2.

## 5.5 Application to Specific Models

We now show that all the results of this chapter apply to the stochastic optimal control problems of Section 2.3.3 and 2.3.4. However, only a portion of the results apply to the minimax control problem of Section 2.3.5, since D.1 is not satisfied in the absence of additional assumptions.

*Stochastic Optimal Control—Outer Integral Formulation*

**Proposition 5.12** Consider the mapping

$$H(x, u, J) = E^*\{g(x, u, w) + \alpha J[f(x, u, w)] | x, u\} \tag{45}$$

of Section 2.3.3 and let $J_0(x) = 0$ for $\forall x \in S$. If

$$0 \le g(x, u, w) \qquad \forall x \in S, \quad u \in U(x), \quad w \in W, \tag{46}$$

then Assumptions I, I.1, and I.2 are satisfied with the scalar in I.2 equal to $\alpha$. If

$$g(x, u, w) \le 0 \qquad \forall x \in S, \quad u \in U(x), \quad w \in W, \tag{47}$$

then Assumptions D, D.1, and D.2 are satisfied with the scalar in D.2 equal to $\alpha$.

*Proof* Assumptions I and D are trivially satisfied in view of (46) or (47), respectively, and the fact that $J_0(x) = 0$ for $\forall x \in S$. Assumptions I.1 and D.1 are satisfied in view of the monotone convergence theorem for outer integration (Proposition A.1). From Lemma A.2 we have under (46) for all

$r > 0$ and $J \in F$ with $J_0 \leq J$

$$\begin{aligned} H(x, u, J + r) &= E^*\{g(x, u, w) + \alpha J[f(x, u, w)] + \alpha r | x, u\} \\ &= E^*\{g(x, u, w) + \alpha J[f(x, u, w)] | x, u\} + \alpha r \\ &= H(x, u, J) + \alpha r. \end{aligned}$$

Hence I.2 is satisfied as stated in the proposition. Under (47), we have from Lemmas A.2 and A.3(c) that for all $r > 0$ and $J \in F$ with $J \leq J_0$

$$H(x, u, J - r) = H(x, u, J) - \alpha r,$$

and D.2 is satisfied. Q.E.D.

Thus all the results of the previous sections apply to stochastic optimal control problems with additive cost functionals. In fact, under additional countability assumptions it is possible to exploit the additive structure of these problems and obtain results relating to the existence of optimal or nearly optimal stationary policies under Assumption D. These results are stated in the following proposition. A proof of part (a) may be found in Blackwell [B10]. Proofs of parts (b) and (c) may be found in Ornstein [O4] and Frid [F2].

**Proposition 5.13** Consider the mapping

$$H(x, u, J) = E\{g(x, u, w) + J[f(x, u, w)] | x, u\}$$

of Section 2.3.2 ($W$ is countable), and let $J_0(x) = 0$ for all $x \in S$. Assume that $S$ is countable, $J^*(x) > -\infty$ for all $x \in S$, and $g$ satisfies

$$b \leq g(x, u, w) \leq 0 \qquad \forall x \in S, \quad u \in U(x), \quad w \in W,$$

where $b \in (-\infty, 0)$ is some scalar. Then:

(a) If for each $x \in S$ there exists a policy which is optimal at $x$, then there exists a stationary optimal policy.

(b) For every $\varepsilon > 0$ there exists a $\mu_\varepsilon \in M$ such that

$$J_{\mu_\varepsilon}(x) \leq (1 - \varepsilon) J^*(x) \qquad \forall x \in S.$$

(c) If there exists a scalar $\beta \in (-\infty, 0)$ such that $\beta \leq J^*(x)$ for all $x \in S$, then for every $\varepsilon > 0$, there exists a stationary $\varepsilon$-optimal policy, i.e., a $\mu_\varepsilon \in M$ such that

$$J^* \leq J_{\mu_\varepsilon} \leq J^* + \varepsilon.$$

We note that the conclusion of part (a) may fail to hold if we have $J^*(x) = -\infty$ for some $x \in S$, even if $S$ is finite, as shown by a counterexample found in Blackwell [B10]. The conclusions of parts (b) and (c) may fail to hold if $S$ is uncountable, as shown by a counterexample due to Ornstein [O4]. The conclusion of part (c) may fail to hold if $J^*$ is unbounded below, as shown by a counterexample due to Blackwell [B8]. We also note that the results of Proposition 5.13 can be strengthened considerably in the special case of a deterministic optimal control problem (cf. the mapping of Section 2.3.1). These results are given in Bertsekas and Shreve [B6].

*Stochastic Optimal Control—Multiplicative Cost Functional*

**Proposition 5.14** Consider the mapping

$$H(x,u,J) = E\{g(x,u,w)J[f(x,u,w)]|x,u\}$$

of Section 2.3.4 and let $J_0(x) = 1$ for $\forall x \in S$.

(a) If there exists a $b \in R$ such that

$$1 \le g(x,u,w) \le b \qquad \forall x \in S, \quad u \in U(x), \quad w \in W,$$

then Assumptions I, I.1, and I.2 are satisfied with the scalar in I.2 equal to $b$.

(b) If

$$0 \le g(x,u,w) \le 1 \qquad \forall x \in S, \quad u \in U(x), \quad w \in W,$$

then Assumptions D, D.1, and D.2 are satisfied with the scalar in D.2 equal to unity.

*Proof* This follows easily from the assumptions and the monotone convergence theorem for ordinary integration. Q.E.D.

*Minimax Control*

**Proposition 5.15** Consider the mapping

$$H(x,u,J) = \sup_{w \in W(x,u)} \{g(x,u,w) + \alpha J[f(x,u,w)]\}$$

of Section 2.3.5 and let $J_0(x) = 0$ for $\forall x \in S$.

(a) If

$$0 \le g(x,u,w) \qquad \forall x \in S, \quad u \in U(x), \quad w \in W,$$

then Assumptions I, I.1, and I.2 are satisfied with the scalar in I.2 equal to $\alpha$.

(b) If

$$g(x,u,w) \le 0 \qquad \forall x \in S, \quad u \in U(x), \quad w \in W,$$

then Assumptions D and D.2 are satisfied with the scalar in D.2 equal to $\alpha$.

*Proof* The proof is entirely similar to the one of Proposition 5.12. Q.E.D.

*Chapter 6*

# A Generalized Abstract Dynamic Programming Model

As we discussed in Section 2.3.2, there are certain difficulties associated with the treatment of stochastic control problems in which the space $W$ of the stochastic parameter is uncountable. For this reason we resorted to outer integration in the model of Section 2.3.3. The alternative explored in this chapter is to modify the entire framework so that policies $\pi = (\mu_0, \mu_1, \ldots)$ consist of functions $\mu_k$ from a strict subset of $M$—for example, those functions which are appropriately measurable. This approach is related to the one we employ in Part II. Unfortunately, however, many of our earlier results and particularly those of Chapter 5 cannot be proved within the generalized framework to be introduced. The results we provide are sufficient, however, for a satisfactory treatment of finite horizon models and infinite horizon models under contraction assumptions. Some of the results of Chapter 5 proved under Assumption D also have counterparts within the generalized framework. The reader, aided by our subsequent discussion, should be able to easily recognize these results.

Certain aspects of the framework of this chapter may seem somewhat artificial to the reader at this point. The motivation for our line of analysis stems primarily from ideas that are developed in Part II, and the reader may wish to return to this chapter after gaining some familiarity with Part II.

The results provided in the following sections are applied to a stochastic optimal control problem with multiplicative cost functional in Section 11.3. The analysis given there illustrates clearly the ideas underlying our development in this chapter.

## 6.1 General Remarks and Assumptions

Consider the sets $S$, $C$, $U(x)$, $M$, $\Pi$, and $F$ introduced in Section 2.1. We consider in addition two subsets $F^*$ and $\tilde{F}$ of the set $F$ of extended real-valued functions on $S$ satisfying

$$F^* \subset \tilde{F} \subset F$$

and a subset $\tilde{M}$ of the set $M$ of functions $\mu: S \to C$ satisfying $\mu(x) \in U(x)$ for $\forall x \in S$. The subset of policies in $\Pi$ corresponding to $\tilde{M}$ is denoted by $\tilde{\Pi}$, i.e.,

$$\tilde{\Pi} = \{(\mu_0, \mu_1, \ldots) \in \Pi \,|\, \mu_k \in \tilde{M},\ k = 0, 1, \ldots\}.$$

In place of the mapping $H$ of Section 2.1, we consider a mapping $H: SC\tilde{F} \to R^*$ satisfying for all $x \in S$, $u \in U(x)$, $J$, $J' \in \tilde{F}$, the *monotonicity assumption*

$$H(x, u, J) \le H(x, u, J') \qquad \text{if} \quad J \le J'.$$

Thus the mapping $H$ in this chapter is of the same nature as the one of Chapters 2–5, the only difference being that $H$ is defined on $SC\tilde{F}$ rather than on $SCF$. Thus if $\tilde{F}$ consists of appropriately measurable functions and $H$ corresponds to a stochastic optimal control problem such as the one of Section 2.3.3 (with $g$ measurable), then $H$ can be defined in terms of ordinary integration rather than outer integration.

For $\mu \in \tilde{M}$ we consider the mapping $T_\mu: \tilde{F} \to F$ defined by

$$T_\mu(J)(x) = H[x, \mu(x), J] \qquad \forall x \in S.$$

Consider also the mapping $T: \tilde{F} \to F$ defined by

$$T(J)(x) = \inf_{u \in U(x)} H(x, u, J) \qquad \forall x \in S.$$

We are given a function $J_0: S \to R^*$ satisfying

$$J_0 \in F^*, \qquad J_0(x) > -\infty \qquad \forall x \in S$$

and we are interested in the $N$-stage problem

$$\begin{aligned} &\text{minimize} \quad J_{N,\pi}(x) = (T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(x) \\ &\text{subject to} \quad \pi \in \tilde{\Pi} \end{aligned} \tag{1}$$

and its infinite horizon counterpart

$$\begin{aligned} &\text{minimize} \quad J_\pi(x) = \lim_{N\to\infty} (T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(x) \\ &\text{subject to} \quad \pi \in \tilde{\Pi}. \end{aligned} \tag{2}$$

We use the notation,

$$J_N^* = \inf_{\pi\in\tilde{\Pi}} J_{N,\pi}, \qquad J^* = \inf_{\pi\in\tilde{\Pi}} J_\pi,$$

and employ the terminology of Chapter 2 regarding optimal, $\varepsilon$-optimal, and stationary policies, as well as sequences of polices exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality.

The following conditions regarding the sets $F^*$, $\tilde{F}$, and $\tilde{M}$ will be assumed in every result of this chapter.

**A.1** For each $x \in S$ and $u \in U(x)$, there exists a $\mu \in \tilde{M}$ such that $\mu(x) = u$. (This implies, in particular, that for every $J \in \tilde{F}$ and $x \in S$

$$T(J)(x) = \inf_{u\in U(x)} H(x, u, J) = \inf_{\mu\in\tilde{M}} H[x, \mu(x), J]).$$

**A.2** For all $J \in F^*$ and $r \in R$, we have

$$T(J) \in F^*, \qquad (J + r) \in F^*.$$

**A.3** For all $J \in \tilde{F}$, $\mu \in \tilde{M}$, and $r \in R$, we have

$$T_\mu(J) \in \tilde{F}, \qquad (J + r) \in \tilde{F}.$$

**A.4** For each $J \in F^*$ and $\varepsilon > 0$, there exists a $\mu_\varepsilon \in \tilde{M}$ such that for all $x \in S$

$$T_{\mu_\varepsilon}(J)(x) \le \begin{cases} T(J)(x) + \varepsilon & \text{if} \quad T(J)(x) > -\infty, \\ -1/\varepsilon & \text{if} \quad T(J)(x) = -\infty. \end{cases}$$

In Section 6.3 the following assumption will also be used.

**A.5** For every sequence $\{J_k\} \subset \tilde{F}$ that converges pointwise, we have $\lim_{k\to\infty} J_k \in \tilde{F}$. If, in addition, $\{J_k\} \subset F^*$, then $\lim_{k\to\infty} J_k \in F^*$.

Note that in the special case where $F^* = \tilde{F} = F$ and $\tilde{M} = M$, we obtain the framework of Chapters 2–5, and all the preceding assumptions are satisfied. Thus this chapter deals with an extension of the framework of Chapters 2–5.

We now provide some examples of sets $F^*$, $\tilde{F}$, and $\tilde{M}$ which are useful in connection with the mapping

$$H(x, u, J) = \int^* \{g(x, u, w) + \alpha J[f(x, u, w)]\} p(dw|x, u)$$

associated with the stochastic optimal control problem of Section 2.3.3. We take $J_0(x) = 0$ for $\forall x \in S$. The terminology employed is explained in Chapter 7.

EXAMPLE 1 Let $S$, $C$, and $W$ be Borel spaces, $\mathscr{F}$ the Borel $\sigma$-algebra on $W$, $f$ a Borel-measurable function mapping $SCW$ into $S$, $g$ a lower semianalytic function mapping $SCW$ into $R^*$, $p(dw|x,u)$ a Borel-measurable stochastic kernel on $W$ given $SC$, and $\alpha$ a positive scalar. Let the set

$$\Gamma = \{(x,u) \in SC | x \in S, u \in U(x)\}$$

be analytic. Take $F^*$ to be the set of extended real-valued, lower semianalytic functions on $S$, $\tilde{F}$ the set of extended real-valued, universally measurable functions on $S$, and $\tilde{M}$ the set of universally measurable mappings from $S$ to $C$ with graph in $\Gamma$ (i.e., $\mu \in \tilde{M}$ if $\mu$ is universally measurable and $(x, \mu(x)) \in \Gamma$ for $\forall x \in S$). This example is the subject of Chapters 8 and 9.

EXAMPLE 2 Same as Example 1 except that $\tilde{M}$ is the set of all analytically measurable mappings from $S$ to $C$ with graph in $\Gamma$. This example is treated in Section 11.2.

EXAMPLE 3 Same as Example 1 except for the following: $p(dw|x,u)$ and $f$ are continuous, $g$ real-valued, upper semicontinuous, and bounded above, $\Gamma$ an open subset of $SC$, $\tilde{F}$ the set of extended, real-valued, Borel-measurable functions on $S$ which are bounded above, $F^*$ the set of extended real-valued, upper semicontinuous functions on $S$ which are bounded above, and $\tilde{M}$ the set of Borel measurable mappings from $S$ to $C$ with graph in $\Gamma$. This is the upper semicontinuous model of Definition 8.8.

EXAMPLE 4 Same as Example 3 except for the following: $C$ is in addition compact, $g$ real-valued, lower semicontinuous, and bounded below, $\Gamma$ a closed subset of $SC$, $\tilde{F}$ the set of extended real-valued, Borel-measurable functions on $S$ which are bounded below, and $F^*$ the set of extended real-valued, lower semicontinuous functions on $S$ which are bounded below. This is a special case of the lower semicontinuous model of Definition 8.7.

All these examples satisfy Assumptions A.1–A.4 stated earlier (see also Sections 7.5 and 7.7). The first two satisfy Assumption A.5 as well.

## 6.2 Analysis of Finite Horizon Models

Simple modifications of some of the assumptions and proofs in Chapter 3 provide a satisfactory analysis of the finite horizon problem (1). We first modify appropriately some of the assumptions of Section 3.1.

**Assumption $\tilde{\mathbf{F}}$.2** Same in statement as Assumption F.2 of Section 3.1 except that $F$ is replaced by $\tilde{F}$.

**Assumption $\tilde{\text{F}}$.3** Same in statement as Assumption F.3 of Section 3.1 except we require that $J \in F^*$, $\{J_n\} \subset \tilde{F}$, and $\{\mu_n\} \subset \tilde{M}$, instead of $J \in F$, $\{J_n\} \subset F$, and $\{\mu_n\} \subset M$.

It can be easily seen that $\tilde{\text{F}}$.2 is satisfied in Examples 1–4 of the previous section. It is also possible to show (see the proof of Proposition 8.4) that $\tilde{\text{F}}$.3 is satisfied in Example 1, where universally measurable policies are employed.

By nearly verbatim repetition of the proofs of Proposition 3.1(b) and Proposition 3.2 we obtain the following.

**Proposition 6.1** (a) Let Assumptions A.1–A.4 and $\tilde{\text{F}}$.2 hold and assume that $J_k^*(x) > -\infty$ for all $x \in S$ and $k = 1, 2, \ldots, N$. Then

$$J_N^* = T^N(J_0),$$

and for every $\varepsilon > 0$ there exists an $N$-stage $\varepsilon$-optimal policy, i.e., a $\pi_\varepsilon \in \tilde{\Pi}$ such that

$$J_N^* \leq J_{N,\pi_\varepsilon} \leq J_N^* + \varepsilon.$$

(b) Let Assumptions A.1–A.4 and $\tilde{\text{F}}$.3 hold and assume that $J_{k,\pi}(x) < \infty$ for all $x \in S$, $\pi \in \tilde{\Pi}$, and $k = 1, 2, \ldots, N$. Then

$$J_N^* = T^N(J_0).$$

Furthermore, given any sequence $\{\varepsilon_n\}$ with $\varepsilon_n \downarrow 0$, $\varepsilon_n > 0$ for $\forall n$, there exists a sequence of policies exhibiting $\{\varepsilon_n\}$-dominated covergence to optimality. In particular, if in addition $J_N^*(x) > -\infty$ for all $x \in S$, then for every $\varepsilon > 0$ there exists an $\varepsilon$-optimal policy.

Similarly, by modifying the proofs of Proposition 3.3 and Corollary 3.3.1(b), we obtain the following.

**Proposition 6.2** Let Assumptions A.1–A.4 hold.

(a) A policy $\pi^* = (\mu_0^*, \mu_1^*, \ldots) \in \tilde{\Pi}$ is uniformly $N$-stage optimal if and only if $(T_{\mu_k^*} T^{N-k-1})(J_0) = T^{N-k}(J_0)$, $k = 0, 1, \ldots, N-1$.

(b) If there exists a uniformly $N$-stage optimal policy, then

$$J_N^* = T^N(J_0).$$

Analogs of Corollary 3.3.1(a) and Proposition 3.4 can be proved if $\tilde{M}$ is rich enough so that the following assumption holds.

**Exact Selection Assumption** For every $J \in F^*$, if the infimum in

$$T(J) = \inf_{u \in U(x)} H(x, u, J)$$

is attained for every $x \in S$, then there exists a $\mu^* \in \tilde{M}$ such that $T_{\mu^*}(J) = T(J)$.

In Examples 1 and 4 of the previous section the exact selection assumption is satisfied (see Propositions 7.50 and 7.33). The following proposition is proved similarly to Corollary 3.3.1(a) and Proposition 3.4.

**Proposition 6.3** Let Assumptions A.1–A.4 and the exact selection assumption hold.

(a) There exists a uniformly $N$-stage optimal policy if and only if the infimum in the relation

$$T^{k+1}(J_0)(x) = \inf_{u \in U(x)} H[x, u, T^k(J_0)]$$

is attained for each $x \in S$ and $k = 0, 1, \ldots, N-1$.

(b) Let the control space $C$ be a Hausdorff space and assume that for each $x \in S$, $\lambda \in R$, and $k = 0, 1, \ldots, N-1$, the set

$$U_k(x, \lambda) = \{u \in U(x) | H[x, u, T^k(J_0)] \leq \lambda\}$$

is compact. Then

$$J_N^* = T^N(J_0),$$

and there exists a uniformly $N$-stage optimal policy.

## 6.3 Analysis of Infinite Horizon Models under a Contraction Assumption

We consider the following modified version of Assumption C of Section 4.1.

**Assumption $\tilde{\mathrm{C}}$** There is a closed subset $\bar{B}$ of the space $B$ such that:

(a) $J_0 \in \bar{B} \cap F^*$,
(b) For all $J \in \bar{B} \cap F^*$, the function $T(J)$ belongs to $\bar{B} \cap F^*$,
(c) For all $J \in \bar{B} \cap \tilde{F}$ and $\mu \in \tilde{M}$, the function $T_\mu(J)$ belongs to $\bar{B} \cap \tilde{F}$.

Furthermore, for every $\pi = (\mu_0, \mu_1, \ldots) \in \tilde{\Pi}$, the limit

$$\lim_{N \to \infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0)(x)$$

exists and is a real number for each $x \in S$. In addition, there exists a positive integer $m$ and scalars $\rho$ and $\alpha$ with $0 < \rho < 1$, $0 < \alpha$ such that

$$\|T_\mu(J) - T_\mu(J')\| \leq \alpha \|J - J'\| \qquad \forall \mu \in \tilde{M}, \quad J, J' \in \bar{B} \cap \tilde{F},$$

$$\|(T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{m-1}})(J) - (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{m-1}})(J')\| \leq \rho \|J - J'\|$$
$$\forall \mu_0, \ldots, \mu_{m-1} \in \tilde{M}, \quad J, J' \in \bar{B} \cap \tilde{F}.$$

If Assumptions A.1–A.5 and $\tilde{C}$ are made, then almost all the results of Chapter 3 have counterparts within our extended framework. The key fact is that, since $\tilde{F}$ and $F^*$ are closed under pointwise limits (Assumption A.5), it follows that $B \cap \tilde{F}$, $\bar{B} \cap \tilde{F}$, $B \cap F^*$, *and* $\bar{B} \cap F^*$ *are closed subsets of* $B$. This is true in view of the fact that convergence of a sequence in $B$ (i.e., in sup norm) implies pointwise convergence. As a result the contraction mapping fixed point theorem can be used in exactly the same manner as in Chapter 3 to establish that, for each $\mu \in \tilde{M}$, $J_\mu$ is the unique fixed point of $T_\mu$ in $\bar{B} \cap \tilde{F}$ and $J^*$ is the unique fixed point of $T$ in $\bar{B} \cap F^*$. Only the modified policy iteration algorithm and the associated Proposition 4.9 have no counterparts in this extended framework. The reason is that our assumptions do not guarantee that Step 3 of the policy iteration algorithm can be carried out. Rather than provide a complete list of the analogs of all propositions in Chapter 4 we state selectively and without proof some of the main results that can be obtained within the extended framework.

**Proposition 6.4** Let Assumptions A.1–A.5 and $\tilde{C}$ hold.

(a) The function $J^*$ belongs to $\bar{B} \cap F^*$ and is the unique fixed point of $T$ within $\bar{B} \cap F^*$. Furthermore, if $J' \in \bar{B} \cap F^*$ is such that $T(J') \le J'$, then $J^* \le J'$, while if $J' \le T(J')$, then $J' \le J^*$.

(b) For every $\mu \in \tilde{M}$, the function $J_\mu$ belongs to $\bar{B} \cap \tilde{F}$ and is the unique fixed point of $T_\mu$ within $\bar{B} \cap \tilde{F}$.

(c) There holds

$$\lim_{N \to \infty} \|T^N(J) - J^*\| = 0 \qquad \forall J \in \bar{B} \cap F^*,$$

$$\lim_{N \to \infty} \|T_\mu^N(J) - J_\mu\| = 0 \qquad \forall J \in \bar{B} \cap \tilde{F}, \quad \mu \in \tilde{M}.$$

(d) A stationary policy $\pi^* = (\mu^*, \mu^*, \ldots) \in \tilde{\Pi}$ is optimal if and only if

$$T_{\mu^*}(J^*) = T(J^*).$$

Equivalently, $\pi^*$ is optimal if and only if $J_{\mu^*} \in \bar{B} \cap F^*$ and

$$T_{\mu^*}(J_{\mu^*}) = T(J_{\mu^*}).$$

(e) For any $\varepsilon > 0$, there exists a stationary $\varepsilon$-optimal policy, i.e., a $\pi_\varepsilon = (\mu_\varepsilon, \mu_\varepsilon, \ldots) \in \tilde{\Pi}$ such that

$$\|J^* - J_{\mu_\varepsilon}\| \le \varepsilon.$$

**Proposition 6.5** Let Assumptions A.1–A.5 and $\tilde{C}$ hold. Assume further that the exact selection assumption of the previous section holds.

(a) If for each $x \in S$ there exists a policy which is optimal at $x$, then there exists an optimal stationary policy.

(b) Let $C$ be a Hausdorff space. If for some $J \in \bar{B} \cap F^*$ and for some positive integer $\bar{k}$ the set

$$U_k(x, \lambda) = \{u \in U(x) | H[x, u, T^k(J)] \leq \lambda\}$$

is compact for all $x \in S$, $\lambda \in R$, and $k \geq \bar{k}$, then there exists an optimal stationary policy.

*Part II*

# Stochastic Optimal Control Theory

*Chapter 7*

# Borel Spaces and Their Probability Measures

This chapter provides the mathematical background required for analysis of the dynamic programming models of the subsequent chapters. The key concept, which is developed in Section 7.3 with the aid of the topological concepts discussed in Section 7.2, is that of a Borel space. In Section 7.4 the set of probability measures on a Borel space is shown to be itself a Borel space, and the relationships between these two spaces are explored. Our general framework for dynamic programming hinges on the properties of analytic sets collected in Section 7.6 and used in Section 7.7 to define and characterize lower semianalytic functions. These functions result from executing the dynamic programming algorithm, so we will want to measurably select at or near their infima to construct optimal or nearly optimal policies. The possibilities for this are also discussed in Section 7.7. A similar analysis in a more specialized case is contained in Section 7.5, which is presented first for pedagogical reasons.

Our presentation is aimed at the reader who is acquainted with the basic notions of topology and measure theory, but is unfamiliar with some of the specialized results relating to separable metric spaces and probability measures on their Borel $\sigma$-algebras.

## 7.1 Notation

We collect here for easy reference many of the symbols used in Part II.

*Operations on Sets*

Let $A$ and $B$ be subsets of a space $X$. The *complement* of $A$ in $X$ is denoted by $A^c$. The *set-theoretic difference* $A - B$ is $A \cap B^c$. We will sometimes write $X - A$ is place of $A^c$. The *symmetric difference* $A \triangle B$ is $(A - B) \cup (B - A)$. If $X$ is a topological space, $\bar{A}$ will denote the *closure* of $A$. If $A_1, A_2, \ldots$ is a sequence of sets such that $A_1 \subset A_2 \subset \cdots$ and $A = \bigcup_{n=1}^{\infty} A_n$, we write $A_n \uparrow A$. If $A_1 \supset A_2 \supset \cdots$ and $A = \bigcap_{n=1}^{\infty} A_n$, we write $A_n \downarrow A$. If $X_1, X_2, \ldots$ is a sequence of spaces, the *Cartesian products* of $X_1, X_2, \ldots, X_n$ and of $X_1, X_2, \ldots$ are denoted by $X_1 X_2 \cdots X_n$ and $X_1 X_2 \cdots$, respectively. If the given spaces have topologies, the product space will have the product topology. Under this topology, convergence in the product space is componentwise convergence in the factor spaces. If the given spaces have $\sigma$-algebras $\mathscr{F}_{X_1}, \mathscr{F}_{X_2}, \ldots$, the product $\sigma$-algebras are denoted by $\mathscr{F}_{X_1}\mathscr{F}_{X_2} \cdots \mathscr{F}_{X_n}$ and $\mathscr{F}_{X_1}\mathscr{F}_{X_2} \cdots$, respectively.

If $X$ and $Y$ are arbitrary spaces and $E \subset XY$, then for each $x \in X$, the *x-section* of $E$ is

$$E_x = \{y \in Y | (x, y) \in E\}. \tag{1}$$

If $\mathscr{P}$ is a class of subsets of a space $X$, we denote by $\sigma(\mathscr{P})$ the smallest $\sigma$-algebra containing $\mathscr{P}$. We denote by $\mathscr{P}_\sigma$ or $\mathscr{P}_\delta$ the class of all subsets which can be obtained by union or intersection, respectively, of countably many sets in $\mathscr{P}$. If $\mathscr{F}$ is the collection of closed subsets of a topological space $X$, then $\mathscr{F}_\delta = \mathscr{F}$, and the members of $\mathscr{F}_\sigma$ are called the *$F_\sigma$-subsets* of $X$. If $\mathscr{G}$ is the collection of open subsets of $X$, the members of $\mathscr{G}_\delta$ are called the *$G_\delta$-subsets* of $X$.

If $(X, \mathscr{P})$ is a paved space, i.e., $\mathscr{P}$ is a nonempty collection of subsets of $X$, and $S$ is a Suslin scheme for $\mathscr{P}$ (Definition 7.15), then $N(S)$ is the nucleus of $S$. The collection of all nuclei of Suslin schemes for $\mathscr{P}$ is denoted $\mathscr{S}(\mathscr{P})$.

*Special Sets*

The symbol $R$ represents the *real numbers* with the usual topology. We use $R^*$ to denote the *extended real numbers* $[-\infty, +\infty]$ with the topology discussed following Definition 7.7 in Section 7.3. Similarly, $Q$ is the set of *rational numbers* and $Q^*$ is the set of extended rational numbers $Q \cup \{\pm\infty\}$.

If $X$ and $Y$ are sets and $f: X \to Y$, the *graph* of $f$ is

$$\operatorname{Gr}(f) = \{(x, f(x)) | x \in X\}. \tag{2}$$

If $A \subset X$ and $\mathscr{C}$ is a collection of subsets of $X$, we define $f(A) = \{f(x)|x \in A\}$ and

$$f(\mathscr{C}) = \{f(C)|C \in \mathscr{C}\}. \tag{3}$$

If $B \subset Y$ and $\mathscr{C}$ is a collection of subsets of $Y$, we define $f^{-1}(B) = \{x \in X|f(x) \in B\}$ and

$$f^{-1}(\mathscr{C}) = \{f^{-1}(C)|C \in \mathscr{C}\}. \tag{4}$$

If $X$ is a topological space, $\mathscr{F}_X$ is the collection of closed subsets of $X$ and $\mathscr{B}_X$ the *Borel σ-algebra* on $X$ (Definition 7.6). The *space of probability measures* on $(X, \mathscr{B}_X)$ is denoted by $P(X)$; $C(X)$ is the *Banach space of bounded, real-valued, continuous functions* on $X$ with the supremum norm

$$\|f\| = \sup_{x \in X} |f(x)|;$$

for any metric $d$ on $X$ which is consistent with its topology, $U_d(X)$ is the *space of bounded real-valued functions on X which are uniformly continuous with respect to d.* If $X$ is a Borel space (Definition 7.7), $\mathscr{A}_X$ is its *analytic σ-algebra* (Definition 7.19) and $\mathscr{U}_X$ its *universal σ-algebra* (Definition 7.18).

We let $N$ denote the *set of positive integers* with the discrete topology. The *Baire null space* $\mathscr{N}$ is the product of countably many copies of $N$. The *Hilbert cube* $\mathscr{H}$ is the product of countably many copies of $[0, 1]$. We will denote by $\Sigma$ the collection of *finite sequences of positive integers.* We impose no topology on $\Sigma$. If $s \in \Sigma$ and $z = (\zeta_1, \zeta_2, \ldots) \in \mathscr{N}$, we write $s < z$ to mean $s = (\zeta_1, \zeta_2, \ldots, \zeta_k)$ for some $k$.

*Mappings*

If $X$ and $Y$ are spaces, $\text{proj}_X$ is the *projection mapping* from $XY$ onto $X$. If $E$ is a subset of $X$, the *indicator function* of $E$ is given by

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases} \tag{5}$$

If $f: X \to [-\infty, +\infty]$, the *positive and negative parts* of $f$ are the functions

$$f^+(x) = \max\{0, f(x)\}, \tag{6}$$

$$f^-(x) = \max\{0, -f(x)\}. \tag{7}$$

If $f_n: X \to Y$ is a sequence of functions, $Y$ is a topological space, and $\lim_{n \to \infty} f_n(x) = f(x)$ for all $x \in X$, then we write $f_n \to f$. If, in addition, $Y = [-\infty, +\infty]$ and $f_1(x) \le f_2(x) \le \cdots$ for all $x \in X$, we write $f_n \uparrow f$, while if $f_1(x) \ge f_2(x) \ge \cdots$ for all $x \in X$, we write $f_n \downarrow f$. In general, when the arguments of extended real-valued functions are omitted, the statements are to be interpreted pointwise. For example, $(\sup_n f_n)(x) = \sup_n f_n(x)$ for all $x \in X$, $f_1 \leq f_2$ if and only if $f_1(x) \leq f_2(x)$ for all $x \in X$, and $f + \varepsilon$ is the function $(f + \varepsilon)(x) = f(x) + \varepsilon$ for all $x \in X$.

*Miscellaneous*

If $(X, d)$ is a nonempty metric space, $x \in X$, and $Y$ is a nonempty subset of $X$, we define the *distance from $x$ to $Y$* by

$$d(x, Y) = \inf_{y \in Y} d(x, y). \tag{8}$$

We define the diameter of $Y$ by

$$\operatorname{diam}(Y) = \sup_{x, y \in Y} d(x, y). \tag{9}$$

If $(X, \mathscr{F})$ is a measurable space and $\mathscr{F}$ contains all singleton sets, then for $x \in X$ we denote by $p_x$ the probability measure on $(X, \mathscr{F})$ which assigns mass one to the set $\{x\}$.

## 7.2 Metrizable Spaces

**Definition 7.1** Let $(X, \mathscr{T})$ be a topological space. A metric $d$ on $X$ is *consistent* with $\mathscr{T}$ if every set of the form $\{y \in X \mid d(x, y) < c\}$, $x \in X$, $c > 0$, is in $\mathscr{T}$, and every nonempty set in $\mathscr{T}$ is the union of sets of this form. The space $(X, \mathscr{T})$ is *metrizable* if such a metric exists.

The distinction between metric and metrizable spaces is a fine one: In a metric space we have settled on a metric, while in a metrizable space the choice is still open. If one metric consistent with the given topology exists, then a multitude of them can be found. For example, if $d$ is a metric on $X$ consistent with $\mathscr{T}$, the metric $\rho$ defined by

$$\rho(x, y) = d(x, y)/[1 + d(x, y)] \qquad \forall x, y \in X$$

is also consistent with $\mathscr{T}$. In what follows, we abbreviate the notation for metrizable spaces, writing $X$ in place of $(X, \mathscr{T})$.

*If $(X, \mathscr{T})$ is a topological space and $Y \subset X$, unless otherwise specified, we will understand $Y$ to be a topological space with open sets $G \cap Y$, where $G$ ranges over $\mathscr{T}$.* This is called the *relative topology*. If $(Z, \mathscr{S})$ is another topological space, $\varphi: Z \to X$ is one-to-one and continuous, and $\varphi^{-1}$ is continuous on $\varphi(Z)$ with the relative topology, we say that $\varphi$ is a *homeomorphism* and $Z$ is *homeomorphic* to $\varphi(Z)$. When there exists a homeomorphism from $Z$ into $X$, we also say that $Z$ can be *homeomorphically embedded* in $X$. Given a metric $d$ on $X$ consistent with its topology and a homeomorphism $\varphi: Z \to X$

as just described, we may define a metric $d_1$ on $Z$ by

$$d_1(z_1, z_2) = d(\varphi(z_1), \varphi(z_2)). \tag{10}$$

It can be easily verified that the metric $d_1$ is consistent with the topology $\mathscr{S}$. This implies that every topological space homeomorphic to a metrizable space (or subset of a metrizable space) is itself metrizable.

Our attention will be focused on metrizable spaces and their Borel $\sigma$-algebras. The presence of a metric in such spaces permits simple proofs of facts whose proofs are quite complicated or even impossible in more general topological spaces. We give two of these as lemmas for later reference.

**Lemma 7.1** (Urysohn's lemma) Let $X$ be a metrizable space and $A$ and $B$ disjoint, nonempty, closed subsets of $X$. Then there exists a continuous function $f: X \to [0, 1]$ such that $f(a) = 0$ for every $a \in A$, $f(b) = 1$ for every $b \in B$, and $0 < f(x) < 1$ for every $x \notin A \cup B$. If $d$ is a metric consistent with the topology on $X$ and $\inf_{a \in A, b \in B} d(a, b) > 0$, then $f$ can be chosen to be uniformly continuous with respect to the metric $d$.

*Proof* Let $d$ be a metric on $X$ consistent with its topology and define

$$f(x) = d(x, A)/[d(x, A) + d(x, B)],$$

where the distance from a point to a nonempty closed set is defined by (8). This distance is zero if and only if the point is in the set, and the mapping of (8) is Lipschitz-continuous by (6) of Appendix C. This $f$ has the required properties. If $\inf_{a \in A, b \in B} d(a, b) > 0$, then $d(x, A) + d(x, B)$ is bounded away from zero, and the uniform continuity of $f$ follows. Q.E.D.

**Lemma 7.2** Let $X$ be a metrizable space. Every closed subset of $X$ is a $G_\delta$ and every open subset is an $F_\sigma$.

*Proof* We prove the first statement; the second follows by complementation. Let $F$ be closed. We may assume without loss of generality that $F$ is nonempty. Let $d$ be a metric on $X$ consistent with its topology. The continuity of the function $x \to d(x, F)$ implies that

$$G_n = \{x \in X | d(x, F) < 1/n\}$$

is open. But $F = \bigcap_{n=1}^{\infty} G_n$. Q.E.D.

**Definition 7.2** Let $X$ be a metrizable topological space. The space $X$ is *separable* if it contains a countable dense set.

It is easily verified that any subspace of a separable metrizable space is separable and metrizable. A collection of subsets of a topological space $(X, \mathscr{T})$ is a *base* for the topology if every open set can be written as a union of sets from the collection. It is a *subbase* if a base can be obtained by taking finite intersections of sets from the collection. If $\mathscr{T}$ has a countable base, $(X, \mathscr{T})$ is said to be *second countable*. A topological space is *Lindelöf* if every collection of open sets which covers the space contains a countable subcollection which also covers the space. It is a standard result that in metrizable spaces, separability, second countability, and the Lindelöf property are equivalent. The following proposition is a direct consequence of this fact.

**Proposition 7.1** Let $(X, \mathscr{T})$ be a separable, metrizable, topological space and $\mathscr{B}$ a base for the topology $\mathscr{T}$. Then $\mathscr{B}$ contains a countable subcollection $\mathscr{B}_0$ which is also a base for $\mathscr{T}$.

*Proof* Let $\mathscr{C}$ be a countable base for the topology $\mathscr{T}$. Every set $C \in \mathscr{C}$ has the form $C = \bigcup_{\alpha \in I(C)} B_\alpha$, where $I(C)$ is an index set and $B_\alpha \in \mathscr{B}$ for every $\alpha \in I(C)$. Since $C$ is Lindelöf, we may assume $I(C)$ is countable. Let $\mathscr{B}_0 = \bigcup_{C \in \mathscr{C}} \{B_\alpha | \alpha \in I(C)\}$. Q.E.D.

The *Hilbert cube* $\mathscr{H}$ is the product of countably many copies of the unit interval (with the product topology). The unit interval is separable and metrizable, and, as we will show later (Proposition 7.4), these properties carry over to the Hilbert cube. In a sense, $\mathscr{H}$ is the canonical separable metrizable space, as the following proposition shows.

**Proposition 7.2** (Urysohn's theorem) Every separable metrizable space is homeomorphic to a subset of the Hilbert cube $\mathscr{H}$.

*Proof* Let $(X, d)$ be a separable metric space with a countable dense set $\{x_k\}$. Define functions

$$\varphi_k(x) = \min\{1, d(x, x_k)\}, \qquad k = 1, 2, \ldots,$$

and $\varphi : X \to \mathscr{H}$ by

$$\varphi(x) = (\varphi_1(x), \varphi_2(x), \ldots).$$

Each $\varphi_k$ is continuous, so $\varphi$ is continuous. (Convergence in $\mathscr{H}$ is componentwise.) If $\varphi(x) = \varphi(y)$, then letting $x_{k_j} \to x$, we see that $\lim_{j \to \infty} d(y, x_{k_j}) = 0$, so $x = y$ and $\varphi$ is one-to-one. It remains to show that $\varphi^{-1}$ is continuous, i.e., $\varphi(y_n) \to \varphi(y)$ implies $y_n \to y$. But if $\varphi(y_n) \to \varphi(y)$, choose $\varepsilon > 0$ and $x_k$ such that $d(y, x_k) < \varepsilon$. Since $d(y_n, x_k) \to d(y, x_k)$ as $n \to \infty$, for $n$ sufficiently large $d(y_n, x_k) < \varepsilon$. Then $d(y, y_n) < 2\varepsilon$. Q.E.D.

If $X$ is a separable metrizable space and $\varphi : X \to \mathscr{H}$ is the homeomorphism whose existence is guaranteed by Proposition 7.2, then by identifying $x \in X$ with $\varphi(x) \in \mathscr{H}$, we can regard $X$ as a subset of $\mathscr{H}$. Indeed, we can regard $X$ as a topological subspace of $\mathscr{H}$, since the images of open sets in $X$ under the mapping $\varphi$ are just the relatively open subsets of $\varphi(X)$ considered as a subspace of $\mathscr{H}$. Note, however, that although $X$ is both open and closed in itself,

$\varphi(X)$ may be neither open nor closed in $\mathscr{H}$. In fact, it may have no topological characterization at all. Likewise, a set with special structure in $X$, say a $G_\delta$, may not have this structure when considered as a subset of $\mathscr{H}$. The next definition and proposition shed some light on this issue.

**Definition 7.3** Let $X$ be a topological space. The space $X$ is *topologically complete* if there is a metric $d$ on $X$ consistent with its topology such that the metric space $(X, d)$ is complete, i.e., if $\{x_k\} \subset X$ is a $d$-Cauchy sequence $[d(x_n, x_m) \to 0$ as $n, m \to \infty]$, then $\{x_k\}$ converges to an element of $X$.

**Proposition 7.3** (Alexandroff's theorem) Let $X$ be a topologically complete space, $Z$ a metrizable space, and $\varphi: X \to Z$ a homeomorphism. Then $\varphi(X)$ is a $G_\delta$-subset of $Z$. Conversely, if $Y$ is a $G_\delta$-subset of $Z$ and $Z$ is topologically complete, then $Y$ is topologically complete.

*Proof* For the proof of the first part of the proposition, we treat $X$ as a subset of $Z$. There are two metrics to consider, a metric $d$ on $Z$ consistent with its topology and a metric $d_1$ on $X$ which makes it complete. Define

$$U_n = \{z \in Z \mid d(z, \bar{X}) < 1/n \text{ and } \exists \text{ an open neighborhood } V(z) \text{ of } z \text{ such that } \sup_{x, y \in V(z) \cap X} d_1(x, y) < 1/n\}.$$

For $n = 1, 2, \ldots$, given $z \in U_n$ and $V(z)$ as just defined, we have

$$V(z) \cap \{y \in Z \mid d(y, \bar{X}) < 1/n\} \subset U_n,$$

so $U_n$ is open. We show $X = \bigcap_{n=1}^{\infty} U_n$.

For $z \in X$, define

$$W(z) = \{y \in X \mid d_1(y, z) < 1/3n\}.$$

Then $W(z)$ is relatively open in $X$, thus of the form $W(z) = V(z) \cap X$, where $V(z)$ is an open neighborhood in $Z$ of $z$. Also,

$$\sup_{x, y \in V(z) \cap X} d_1(x, y) < 1/n,$$

so $z \in U_n$. Therefore $X \subset \bigcap_{n=1}^{\infty} U_n$. Now suppose $z \in \bigcap_{n=1}^{\infty} U_n$. Then $d(z, \bar{X}) = 0$, and since $\bar{X}$ is closed, we have $z \in \bar{X}$. There is a sequence $\{x_k\} \subset X$ such that $x_k \to z$. Let $V_n(z)$ be an open neighborhood in $Z$ of $z$ for which

$$\sup_{x, y \in V_n(z) \cap X} d_1(x, y) < 1/n. \tag{11}$$

For each $n$, there is an index $k_n$ such that $x_k \in V_n(z)$ for $k \geq k_n$. From (11) we see that $d_1(x_i, x_j) < 1/n$ for $i, j \geq k_n$, so $\{x_k\}$ is Cauchy in the complete space $(X, d_1)$ and hence has a limit in $X$. But the limit is $z$ by assumption, so $X = \bigcap_{n=1}^{\infty} U_n$.

For the converse part of the theorem, suppose $(Z, d)$ is a complete metric space and $Y = \bigcap_{n=1}^{\infty} U_n$, where each $U_n$ is open in $Z$. Define a metric $d_1$ on $Y$ by

$$d_1(y,z) = d(y,z) + \sum_{n=1}^{\infty} \min\{1/2^n, |[1/d(y, Z - U_n)] - [1/d(z, Z - U_n)]|\}.$$

If $\{y_k\}$ is Cauchy in $(Y, d_1)$, then it is also Cauchy in $(Z, d)$, and thus has a limit $y \in Z$. For each $n$,

$$|[1/d(y_i, Z - U_n)] - [1/d(y_j, Z - U_n)]| \to 0$$

as $i, j \to \infty$, so $[1/d(y_k, Z - U_n)]$ remains bounded as $k \to \infty$. It follows that $y \in U_n$ for every $n$, hence $y \in Y$. Q.E.D.

As we remarked earlier without proof, the Hilbert cube inherits metrizability and separability from the unit interval. It also inherits topological completeness. This is a special case of the fact, which we now prove, that completeness and separability of metrizable spaces are preserved when taking countable products.

**Proposition 7.4** Let $X_1, X_2, \ldots$ be a sequence of metrizable spaces and $Y_n = X_1 X_2 \cdots X_n$, $Y = X_1 X_2 \cdots$. Then $Y$ and each $Y_n$ is metrizable. If each $X_k$ is separable or topologically complete, then $Y$ and each $Y_n$ is separable or topologically complete, respectively.

*Proof* If $d_k$ is a metric on $X_k$ consistent with its topology, then

$$d(y, \hat{y}) = \sum_{k=1}^{\infty} \min\{1/2^k, d_k(\eta_k, \hat{\eta}_k)\},$$

where $y = (\eta_1, \eta_2, \ldots)$, $\hat{y} = (\hat{\eta}_1, \hat{\eta}_2, \ldots)$, is a metric on $Y$ consistent with the product topology. If each $(X_k, d_k)$ is complete, clearly $(Y, d)$ is complete. If $\mathscr{G}_k$ is a countable base for the topology on $X_k$, the collection of sets of the form $G_1 G_2 \cdots G_n X_{n+1} X_{n+2} \cdots$, where $G_k$ ranges over $\mathscr{G}_k$ and $n$ ranges over the positive integers, is a countable base for the product topology on $Y$. The arguments for the product spaces $Y_n$ are similar. Q.E.D.

Combining Propositions 7.2–7.4, we see that every separable, topologically complete space is homeomorphic to a $G_\delta$-subset of the Hilbert cube, and conversely, every $G_\delta$-subset of the Hilbert cube is separable and topologically complete. We state a second consequence of these propositions as a corollary.

**Corollary 7.4.1** Every separable, topologically complete space can be homeomorphically embedded as a dense $G_\delta$-set in a compact metric space.

*Proof* Let $X$ be separable and topologically complete and let $\varphi$ be a homeomorphism from $X$ into $\mathscr{H}$. Since $\mathscr{H}$ is metrizable, $\varphi(X)$ is a $G_\delta$-subset of $\mathscr{H}$ (Proposition 7.3) and thus a dense $G_\delta$-subset of $\overline{\varphi(X)}$. Tychonoff's theorem implies that $\mathscr{H}$ is compact, so $\overline{\varphi(X)}$ is compact. Q.E.D.

If $X$ and $Z$ are topological spaces, $\varphi$ a homeomorphism from $Z$ *onto* $X$, and $d$ a metric on $X$ consistent with its topology such that $(X, d)$ is complete, then $d_1$ defined by (10) is a metric on $Z$ consistent with its topology, and $(Z, d_1)$ is also complete. Thus topological completeness is preserved under homeomorphisms. The same is true for separability, as is well known. Topological completeness is somewhat different from separability, however, in that one must produce a metric to verify it. It is quite possible that a space has two metrics consistent with its topology, is a complete metric space with one, but is not a complete metric space with the other. For example, let $X = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$ have the discrete topology,

$$d_1(x, y) = \begin{cases} 0 & \text{if} \quad x = y, \\ 1 & \text{if} \quad x \neq y, \end{cases}$$

and

$$d_2(x, y) = |x - y|.$$

Then $(X, d_1)$ is complete, but $(X, d_2)$ is not. A more surprising example is that the set $\mathscr{N}_0$ of irrational numbers between 0 and 1 with the usual topology is topologically complete. To see this, write $\mathscr{N}_0 = \bigcap_{r \in Q}([0, 1] - \{r\})$, where $Q$ is the set of rational numbers. It follows that $\mathscr{N}_0$ is a $G_\delta$-subset of $[0, 1]$ and is thus topologically complete by Proposition 7.3. Another proof is obtained as follows. Let $N$ be the set of positive integers with the discrete topology and $\mathscr{N}$ the product of countably many copies of $N$. The space $\mathscr{N}$ is called the *Baire null space* and is topologically complete (Proposition 7.4). The topological completeness of $\mathscr{N}_0$ follows from the fact that $\mathscr{N}$ and $\mathscr{N}_0$ are homeomorphic. We give the rather lengthy proof of this because it is not readily available elsewhere. This homeomorphism will be used only to construct a counterexample (Example 1 in Chapter 8), so it may be skipped by the reader without loss of continuity.

**Proposition 7.5** The topological spaces $\mathscr{N}_0$ and $\mathscr{N}$ are homeomorphic.

*Proof* Let $\Sigma$ be the set of finite sequences of positive integers. If $z \in \Sigma \cup \mathscr{N}$, we will represent its components by $\zeta_k$. Similarly, $\hat{\zeta}_k$ will represent the components of an element $\hat{z}$ of $\Sigma \cup \mathscr{N}$. The *length* of $z \in \Sigma \cup \mathscr{N}$ is defined to be the number of its components. If $z$ has length greater than or equal to $k$, we define $z_k = (\zeta_k, \zeta_{k+1}, \ldots)$ or $z_k = (\zeta_k, \ldots, \zeta_m)$, depending on whether $z$ has infinite length or length $m < \infty$.

For $z \in \Sigma \cup \mathcal{N}$, define a sequence whose initial terms are

$$x_1(z) = \zeta_1^{-1},$$
$$x_2(z) = (\zeta_1 + \zeta_2^{-1})^{-1},$$
$$x_3(z) = (\zeta_1 + (\zeta_2 + \zeta_3^{-1})^{-1})^{-1}.$$

If $z$ has length $k < \infty$, we define $x_1(z), x_2(z), \ldots, x_k(z)$ as shown, and $x_{k+j}(z) = x_k(z), j = 1, 2, \ldots$.

*Claim 1* The sequence $\{x_k(z)\}$ converges to an element of $(0,1]$ for $\forall z \in \Sigma \cup \mathcal{N}$.

If $z$ has finite length, the claim is trivial. If $z$ has infinite length, then

$$0 < x_{2n}(z) < x_{2n+2}(z) < x_{2n+1}(z) < x_{2n-1}(z) \leq 1, \qquad n = 1, 2, \ldots, \tag{12}$$

so for every $n$

$$x_{2n}(z) \leq \liminf_{k \to \infty} x_k(z) \leq \limsup_{k \to \infty} x_k(z) \leq x_{2n-1}(z). \tag{13}$$

Now

$$\begin{aligned}
0 &< x_{2n-1}(z) - x_{2n}(z) \\
&= [\zeta_1 + x_{2n-2}(z_2)]^{-1} - [\zeta_1 + x_{2n-1}(z_2)]^{-1} \\
&= [\zeta_1 + x_{2n-2}(z_2)]^{-1}[\zeta_1 + x_{2n-1}(z_2)]^{-1}[x_{2n-1}(z_2) - x_{2n-2}(z_2)] \\
&\leq [\zeta_1 + (\zeta_2 + 1)^{-1}]^{-2}[x_{2n-1}(z_2) - x_{2n-2}(z_2)] \\
&\leq [\zeta_1 + (\zeta_2 + 1)^{-1}]^{-2}[\zeta_2 + (\zeta_3 + 1)^{-1}]^{-2}[x_{2n-3}(z_3) - x_{2n-2}(z_3)] \\
&\ \vdots \\
&\leq [\zeta_1 + (\zeta_2 + 1)^{-1}]^{-2}[\zeta_2 + (\zeta_3 + 1)^{-1}]^{-2} \cdots [\zeta_{2n-2} + (\zeta_{2n-1} + 1)^{-1}]^{-2}.
\end{aligned}$$

Since

$$[\zeta_{2k-1} + (\zeta_{2k} + 1)^{-1}]^{-2}[\zeta_{2k} + (\zeta_{2k+1} + 1)^{-1}]^{-2} \leq \tfrac{4}{9}, \qquad k = 1, \ldots, n-1,$$

we have

$$0 < x_{2n-1}(z) - x_{2n}(z) \leq (\tfrac{4}{9})^{n-1},$$

and Claim 1 follows.

Define $\varphi: \Sigma \cup \mathcal{N} \to (0,1]$ by

$$\varphi(z) = \lim_{k \to \infty} x_k(z).$$

Note that if $z \in \mathcal{N}$, then $0 < \varphi(z) < 1$. Also, if $z$ has length at least $k$, then

$$\varphi(z) = \varphi[(\zeta_1, \zeta_2, \ldots, \zeta_{k-1}, 1/\varphi(z_k))]. \tag{14}$$

*Claim 2* If $z \in \mathcal{N}$ and $\varphi(z) = \varphi(\hat{z})$, then $z = \hat{z}$.

Suppose $\varphi(z) = \varphi(\hat{z})$ and $z \neq \hat{z}$. We can use (14) to assume without loss of generality that $\zeta_1 \neq \hat{\zeta}_1$ or else $\hat{z}$ has length one and $\zeta_1 = \hat{\zeta}_1$. In the latter case, (12) implies

$$\varphi(\hat{z}) = 1/\hat{\zeta}_1 = 1/\zeta_1 = x_1(z) > x_3(z) \geq \varphi(z),$$

and a contradiction is reached. In the former case, if $\hat{z}$ has length one, then from (14)

$$1/\hat{\zeta}_1 = \varphi(\hat{z}) = \varphi(z) = 1/[\zeta_1 + \varphi(z_2)],$$

so

$$\hat{\zeta}_1 = \zeta_1 + \varphi(z_2),$$

which is impossible, since $0 < \varphi(z_2) < 1$. If $\hat{z}$ has length greater than one, then

$$1/[\hat{\zeta}_1 + \varphi(\hat{z}_2)] = \varphi(\hat{z}) = \varphi(z) = 1/[\zeta_1 + \varphi(z_2)],$$

and

$$\hat{\zeta}_1 + \varphi(\hat{z}_2) = \zeta_1 + \varphi(z_2).$$

This is also impossible, since $0 < \varphi(\hat{z}_2) \leq 1$ and $0 < \varphi(z_2) < 1$.

*Claim 3* Every rational number in $(0,1]$ has the form $\varphi(\hat{z})$, where $\hat{z} \in \Sigma$.

Let $r_1/q$ be a rational number in $(0,1]$ reduced to lowest terms, $r_1$ and $q$ positive integers. Then

$$r_1/q = (q/r_1)^{-1} = [q_1 + (r_2/r_1)]^{-1},$$

where $q_1$ and $r_2$ are positive integers and $r_2 < r_1$. Likewise,

$$r_2/r_1 = (r_1/r_2)^{-1} = [q_2 + (r_3/r_2)]^{-1},$$

where $q_2$ and $r_3$ are positive integers and $r_3 < r_2$. Continuing, we eventually obtain $r_n = 1$ and have

$$r_1/q = \varphi[(q_1, q_2, \ldots, q_{n-1}, r_{n-1})].$$

Claims 2 and 3 imply that if $z \in \mathcal{N}$, then $\varphi(z)$ is irrational. Put another way, $\varphi$ maps $\mathcal{N}$ into $\mathcal{N}_0$. But given $y \in \mathcal{N}_0$, it is possible to choose positive integers $\zeta_1, \zeta_2, \ldots$, such that

$$(\zeta_1 + 1)^{-1} < y < \zeta_1^{-1},$$
$$(\zeta_1 + \zeta_2^{-1})^{-1} < y < (\zeta_1 + (\zeta_2 + 1)^{-1})^{-1},$$
$$(\zeta_1 + (\zeta_2 + (\zeta_3 + 1)^{-1})^{-1})^{-1} < y < (\zeta_1 + (\zeta_2 + \zeta_3^{-1})^{-1})^{-1},$$

etc., so that defining $z = (\zeta_1, \zeta_2, \ldots)$, we have

$$x_{2k}(z) < y < x_{2k-1}(z), \qquad k = 1, 2, \ldots.$$

It follows that $\varphi(z) = y$, so $\varphi$ maps $\mathcal{N}$ onto $\mathcal{N}_0$ and, by Claim 2, is one-to-one on $\mathcal{N}$.

We show that $\varphi$ restricted to $\mathcal{N}$ is open and continuous. Let $V \subset \mathcal{N}$ be open. We may assume without loss of generality that

$$V = \{z \in \mathcal{N} | (\zeta_1, \ldots, \zeta_n) = (\hat{\zeta}_1, \ldots, \hat{\zeta}_n)\}.$$

Then

$$\varphi(V) = \{(\hat{\zeta}_1 + (\hat{\zeta}_2 + \cdots + (\hat{\zeta}_n + \varphi(z))^{-1} \cdots)^{-1})^{-1} | z \in \mathcal{N}\},$$

and since $\{\varphi(z) | z \in \mathcal{N}\} = \mathcal{N}_0, \varphi(V)$ is open. Since convergence in $\mathcal{N}$ is componentwise and $x_n(z)$ depends only on the first $n$ components of $z \in \mathcal{N}$, continuity of $\varphi$ on $\mathcal{N}$ follows from (13). Q.E.D.

We now examine properties of metrizable spaces related to the notion of total boundedness.

**Definition 7.4** A metric space $(X, d)$ is *totally bounded* if, given $\varepsilon > 0$, there exists a finite subset $F_\varepsilon$ of $X$ for which

$$X = \bigcup_{x \in F_\varepsilon} \{y \in X | d(x, y) < \varepsilon\}.$$

A totally bounded metric space is necessarily separable, since $\bigcup_{n=1}^{\infty} F_{1/n}$ is a countable dense subset. Total boundedness depends on the metric, however, and a space which is totally bounded (and separable) with one metric may not be totally bounded with another. Like separability, total boundedness is preserved under passage to subspaces, i.e., if $(X, d)$ is totally bounded and $Y \subset X$, then $(Y, d)$ is totally bounded. To see this, take $\varepsilon > 0$ and let $F_{\varepsilon/2}$ be a finite subset of $X$ such that

$$X = \bigcup_{x \in F_{\varepsilon/2}} \{y \in X | d(x, y) < \varepsilon/2\}.$$

Choose a point, if possible, in each of the sets

$$Y \cap \{y \in X | d(x, y) < \varepsilon/2\}, \qquad x \in F_{\varepsilon/2},$$

and call the collection of these points $G_\varepsilon$. Then

$$Y = \bigcup_{y \in G_\varepsilon} \{z \in Y | d(y, z) < \varepsilon\}.$$

We use this fact to prove the following classical result relating completeness, compactness, and total boundedness.

**Proposition 7.6** A metric space is compact if and only if it is complete and totally bounded.

*Proof* If $(X, d)$ is a compact metric space, then every Cauchy sequence has an accumulation point. The Cauchy property implies that the sequence converges to this point, and completeness follows. Also, for $\varepsilon > 0$, the collection of sets

$$\{y \in X \mid d(x, y) < \varepsilon\}, \qquad x \in X,$$

contains a finite cover of $X$. Hence, $(X, d)$ is totally bounded.

If $(X, d)$ is complete and totally bounded and $S = \{s_j\}$ is a sequence in $(X, d)$, then an infinite subsequence $S_1 \subset S$ must lie in some set $B_1 = \{y \in X \mid d(x_1, y) < 1\}$. Since $B_1$ is totally bounded, an infinite subsequence $S_2 \subset S_1$ must lie in some set $B_2 = \{y \in B_1 \mid d(x_2, y) < \frac{1}{2}\}$. Continuing in this manner, we have for each $n$ an infinite sequence $S_{n+1} \subset S_n$ lying in $B_{n+1} = \{y \in B_n \mid d(x_{n+1}, y) < 1/(n+1)\}$. Let $j_1 < j_2 < \cdots$ be such that $s_{j_n} \in S_n$. Then $\{s_{j_n}\}$ is Cauchy and thus convergent. Therefore $S$ has an accumulation point, and the compactness of $(X, d)$ follows. Q.E.D.

**Corollary 7.6.1** The Hilbert cube is totally bounded under any metric consistent with its topology, and every separable metrizable space has a totally bounded metrization.

*Proof* The Hilbert cube is compact by Tychonoff's theorem. Urysohn's theorem (Proposition 7.2) can be used to homeomorphically embed a given separable metrizable space into the Hilbert cube. Q.E.D.

As mentioned previously, total boundedness implies separability. By combining this fact with Proposition 7.6, we obtain the following corollary.

**Corollary 7.6.2** A compact metric space is complete and separable.

If $X$ is a metrizable space, the set of all bounded, continuous, real-valued functions on $X$ is denoted $C(X)$. As is well known, $C(X)$ is a Banach space under the norm

$$\|f\| = \sup_{x \in X} |f(x)|,$$

and we will always take $C(X)$ to have the metric and topology corresponding to this norm. If $d$ is a metric on $X$ consistent with its topology, we denote by $U_d(X)$ the collection of functions in $C(X)$ which are uniformly continuous with respect to $d$. We take $U_d(X)$ to have the relative topology of $C(X)$. We conclude this section with a discussion of the properties $C(X)$ and $U_d(X)$ inherit from $X$.

**Proposition 7.7** If $X$ is a compact metrizable space, then $C(X)$ is separable.

*Proof* The space $X$ is separable (Corollary 7.6.2). Let $\{x_k\}$ be a countable dense subset of $X$ and let $F_1, F_2, \ldots$ be an enumeration of the collection of sets of the form $\{y \in X \mid d(x_k, y) \le 1/n\}$, where $k$ and $n$ range over the positive integers. For any disjoint pair $F_i$ and $F_j$, let $f_{ij}$ be a continuous function taking values in $[0,1]$ such that $f_{ij}(x) = 0$ for $x \in F_i$ and $f_{ij}(x) = 1$ for $x \in F_j$. If $F_i$ and $F_j$ are not disjoint, let $f_{ij}$ be identically one. Let $\mathscr{C}$ consist of the functions $f_{ij}$ as $i$ and $j$ range over the positive integers. The collection $\mathscr{C}$ clearly separates points in $X$, i.e., given $x \neq y$, there exists $f \in \mathscr{C}$ for which $f(x) \neq f(y)$. Let $\mathscr{P}$ be the collection of finite-degree polynomials over $\mathscr{C}$, i.e., a typical element in $\mathscr{P}$ has the form

$$\sum_{(i_1,\ldots,i_n),(j_1,\ldots,j_n)} \alpha(i_1,\ldots,i_n;j_1,\ldots,j_n) f_{i_1}^{j_1}\cdots f_{i_n}^{j_n},$$

where $\alpha(i_1,\ldots,i_n;j_1,\ldots,j_n) \in R$, $f_{i_1},\ldots,f_{i_n} \in \mathscr{C}$, and the summation is finite. Then $\mathscr{P}$ is a vector space under addition and the product of two elements in $\mathscr{P}$ is again in $\mathscr{P}$. With these operations $\mathscr{P}$ is an *algebra*, and by the Stone–Weierstrass theorem, $\mathscr{P}$ is dense in $C(X)$. Let $\mathscr{P}_0$ be the collection of finite-degree polynomials over $\mathscr{C}$ with rational coefficients. An easy approximation argument shows that $\mathscr{P}_0$ is dense in $\mathscr{P}$, and thus dense in $C(X)$ as well. Since $\mathscr{P}_0$ is countable, $C(X)$ is separable. Q.E.D.

**Definition 7.5** Let $(X,d_1)$ and $(Y,d_2)$ be metric spaces. A mapping $\varphi: X \to Y$ is an *isometry* if

$$d_1(x_1,x_2) = d_2(\varphi(x_1),\varphi(x_2)) \qquad \forall x_1,x_2 \in X.$$

In this case we say that $(X,d_1)$ and $(\varphi(X),d_2)$ are *isometric* spaces.

If $(X,d_1)$ and $(Y,d_2)$ are as in Definition 7.5, we may regard the former as a subspace of the later, and the distances between points in $X$ are unaffected by this embedding. Thus an isometry is a metric-preserving homeomorphism.

**Proposition 7.8** Let $(X,d)$ be a metric space. There exists a complete metric space $(X_d,d_1)$, called the *completion of* $(X,d)$, and an isometry $\varphi: X \to X_d$ such that $\varphi(X)$ is dense in $X_d$.

*Proof* The construction of the completion of a metric space is standard, so we content ourselves with a sketch of it. Given the metric space $(X,d)$, define an equivalence relation $\sim$ on the set of Cauchy sequences in $(X,d)$ by

$$\{x_n\} \sim \{x_n'\} \Leftrightarrow \lim_{n\to\infty} d(x_n,x_n') = 0.$$

Let $X_d$ be the set of equivalence classes of Cauchy sequences in $(X,d)$ under this relation and let $d_1$ be defined on $X_d X_d$ by

$$d_1(x,y) = \lim_{n\to\infty} d(x_n,y_n), \tag{15}$$

where $\{x_n\}$ and $\{y_n\}$ are chosen to represent the equivalence classes $x$ and $y$. It is straightforward to verify that the limit in (15) exists for every pair of Cauchy sequences $\{x_n\}$ and $\{y_n\}$, and it is independent of the particular sequences chosen to represent the equivalence classes $x$ and $y$. Furthermore, $(X_d, d_1)$ can be shown to be a complete metric space, and the mapping $\varphi$ which takes $x \in X$ into the equivalence class in $X_d$ containing the Cauchy sequence $(x, x, \ldots)$ is an isometry. The image of $X$ under $\varphi$ is dense in $X_d$.
Q.E.D.

We can regard $X_d$ as consisting of $X$ together with limits of all Cauchy sequences in $X$. We are really interested in the case in which $(X, d)$ is totally bounded, for which we have the following result.

**Corollary 7.8.1** Let $(X, d)$ be a totally bounded metric space. There exists a compact metric space $(X_d, d_1)$ and an isometry $\varphi: X \to X_d$ such that $\varphi(X)$ is dense in $X_d$.

*Proof* In light of Propositions 7.6 and 7.8, it suffices to prove that the completion $(X_d, d_1)$ of $(X, d)$ is totally bounded. Choose $\varepsilon > 0$. Regarding $(X, d)$ as a subspace of $(X_d, d_1)$, choose a finite set $F_\varepsilon$ of $X$ for which

$$X = \bigcup_{x \in F_\varepsilon} \{y \in X \mid d(x, y) < \varepsilon/2\}.$$

Since $X$ is dense in $X_d$, we have

$$X_d = \bigcup_{x \in F_\varepsilon} \{y \in X_d \mid d_1(x, y) < \varepsilon\}. \qquad \text{Q.E.D.}$$

If $X$ is a separable metrizable space, it is not necessarily true that $C(X)$ is separable (unless $X$ is compact, in which case we have Proposition 7.7). For example, let $f: R \to [0, 1]$ be defined as

$$f(x) = \begin{cases} 0 & \text{if } |x| \geq \frac{1}{2}, \\ 1 + 2x & \text{if } -\frac{1}{2} \leq x \leq 0, \\ 1 - 2x & \text{if } 0 \leq x \leq \frac{1}{2}, \end{cases}$$

and given an infinite sequence $b = (\beta_1, \beta_2, \ldots)$ of zeroes and ones, define

$$f_b(x) = \sum_{\{n \mid \beta_n = 1\}} f(x - n).$$

We have constructed an uncountable collection of functions $f_b$ in $C(R)$ such that if $b_1 \neq b_2$, then $\|f_{b_1} - f_{b_2}\| = 1$. Therefore, $C(R)$ cannot be separable.

It is true, however, that given a separable metrizable space $X$, there is a metric $d$ on $X$ consistent with its topology such that $U_d(X)$ is separable. This is a consequence of the next proposition and the fact that separability implies the existence of a totally bounded metrization (Corollary 7.6.1). We prove this proposition with the aid of the following lemma.

**Lemma 7.3** Let $Y$ be a metrizable space, $d$ a metric on $Y$ consistent with its topology, and $X \subset Y$. If $g \in U_d(X)$, then $g$ has a continuous extension to $Y$, i.e., there exists $\hat{g} \in C(Y)$ such that $g(x) = \hat{g}(x)$ for every $x \in X$, and the extension $\hat{g}$ can be chosen to satisfy $\|g\| = \|\hat{g}\|$. If $X$ is dense in $Y$, $\hat{g}$ is unique.

*Proof* Since $g$ is uniformly continuous on $X$, given $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ such that if $x_1, x_2 \in X$ and $d(x_1, x_2) \leq \delta(\varepsilon)$, then $|g(x_1) - g(x_2)| \leq \varepsilon$. Suppose $y \in \bar{X}$. Then there exists a sequence $\{x_n\} \subset X$ for which $x_n \to y$. Given $\varepsilon > 0$, there exists $N(\varepsilon)$ such that $d(x_n, x_m) \leq \delta(\varepsilon)$ for all $n$, $m \geq N(\varepsilon)$, so $\{g(x_n)\}$ is Cauchy in $R$. Define $\hat{g}(y) = \lim_{n\to\infty} g(x_n)$. Note that $n \geq N(\varepsilon)$ implies $|g(x_n) - \hat{g}(y)| \leq \varepsilon$.

Suppose now that $x \in X$ and $d(x, y) \leq \delta(\varepsilon)/2$. Choose $n \geq N(\varepsilon)$ so that $d(x_n, y) \leq \delta(\varepsilon)/2$. Then $d(x, x_n) \leq \delta(\varepsilon)$ and

$$|g(x) - \hat{g}(y)| \leq |g(x) - g(x_n)| + |g(x_n) - \hat{g}(y)| \leq 2\varepsilon. \tag{16}$$

This shows that for any sequence $\{x_n'\} \subset X$ with $x_n' \to y$, we have $\hat{g}(y) = \lim_{n\to\infty} g(x_n')$, so the definition of $\hat{g}(y)$ is independent of the particular sequence $\{x_n\}$ chosen. If $y \in X$, we can take $x_n = y$, $n = 1, 2, \ldots$, and obtain $g(y) = \hat{g}(y)$, so $\hat{g}$ is an extension of $g$. If $\{y_m\}$ is a sequence in $\bar{X}$ which converges to $y \in \bar{X}$, then there exist sequences $\{x_{mn}\}$ in $X$ with $y_m = \lim_{n\to\infty} x_{mn}$. Choose $n_1 < n_2 < \cdots$ so that $\lim_{m\to\infty} x_{mn_m} = y$ and $d(x_{mn_m}, y_m) \leq \delta(1/m)/2$. Then

$$\hat{g}(y) = \lim_{m\to\infty} g(x_{mn_m}), \tag{17}$$

and, by (16),

$$|g(x_{mn_m}) - \hat{g}(y_m)| \leq 2/m. \tag{18}$$

Letting $m \to \infty$ in (18) and using (17), we conclude that $\hat{g}(y) = \lim_{m\to\infty} \hat{g}(y_m)$ and $\hat{g}$ is continuous on $\bar{X}$. It is clear that

$$\sup_{x\in X} |g(x)| = \sup_{y\in \bar{X}} |\hat{g}(y)|.$$

If $\bar{X} = Y$, $\hat{g}$ is clearly unique and we are done. If $\bar{X}$ is a proper subset of $Y$, use the Tietze extension theorem (see, e.g., Ash [A1] or Dugundji [D7]) to extend $\hat{g}$ to all of $Y$ so that

$$\|g\| = \sup_{y\in Y} |\hat{g}(y)|. \qquad \text{Q.E.D.}$$

**Proposition 7.9** If $(X, d)$ is a totally bounded metric space, then $U_d(X)$ is separable.

*Proof* Corollary 7.8.1 tells us that $(X, d)$ can be isometrically embedded as a dense subset of a compact metric space $(X_d, d_1)$. We regard $X$ as a subspace of $X_d$. Given any $g \in U_d(X)$, by Lemma 7.3, $g$ has a unique extension $\hat{g} \in C(X_d)$ such that $\|g\| = \|\hat{g}\|$. The mapping $g \to \hat{g}$ is linear and norm-preserving, thus an isometry from $U_d(X)$ to $C(X_d)$. The latter space is separable by Proposition 7.7, and the separability of $U_d(X)$ follows. Q.E.D.

## 7.3 Borel Spaces

The constructions necessary for the subsequent theory of dynamic programming are impossible when the state space and control space are arbitrary sets or even when they are arbitrary measurable spaces. For this reason, we introduce the concept of a Borel space, and in this and subsequent sections we develop the properties of Borel spaces which permit these constructions.

**Definition 7.6** If $X$ is a topological space, the smallest $\sigma$-algebra of subsets of $X$ which contains all open subsets of $X$ is called the *Borel $\sigma$-algebra* and is denoted by $\mathscr{B}_X$. The members of $\mathscr{B}_X$ are called the *Borel subsets* of $X$.

If $X$ is separable and metrizable and $\mathscr{F}$ is a $\sigma$-algebra on $X$ containing a subbase $\mathscr{S}$ for its topology, then $\mathscr{F}$ contains $\mathscr{B}_X$. This is because, from Proposition 7.1, any open set in $X$ can be written as a *countable* union of finite intersections of sets in $\mathscr{S}$. Thus we have $\mathscr{B}_X = \sigma(\mathscr{S})$ for any subbase $\mathscr{S}$.

We will often refer to the smallest $\sigma$-algebra containing a class of subsets as the *$\sigma$-algebra generated by the class.* Thus, $\mathscr{B}_X$ is the $\sigma$-algebra generated by the class of open subsets of $X$. Note that $\mathscr{B}_R$ is the class of Borel subsets of the real numbers in the usual sense, i.e., the $\sigma$-algebra generated by the intervals.

Given a class of real-valued functions on a topological space $X$, it is common to speak of the weakest topology with respect to which all functions in the class are continuous. In a similar vein, one can speak of the smallest $\sigma$-algebra with respect to which all functions in the class are measurable. If $X$ is a metrizable space, it is easy to show that its topology is the weakest with respect to which all functions in $C(X)$ are continuous. The following proposition is the analogous result for $\mathscr{B}_X$. In the proof and in subsequent proofs, we will use the fact that for any two sets $\Omega$, $\Omega'$, any collection $\mathscr{C}$ of subsets of $\Omega'$, and any function $f:\Omega \to \Omega'$, we have

$$\sigma[f^{-1}(\mathscr{C})] = f^{-1}[\sigma(\mathscr{C})].$$

**Proposition 7.10** Let $X$ be a metrizable space. Then $\mathscr{B}_X$ is the smallest $\sigma$-algebra with respect to which every function in $C(X)$ is measurable, i.e., $\mathscr{B}_X = \sigma[\bigcup_{f \in C(X)} f^{-1}(\mathscr{B}_R)]$.

*Proof* Denote $\mathscr{F} = \sigma[\bigcup_{f\in C(X)} f^{-1}(\mathscr{B}_R)]$ and let $\mathscr{T}_R$ be the topology of $R$. We have

$$\begin{aligned}\mathscr{F} &= \sigma\left[\bigcup_{f\in C(X)} f^{-1}[\sigma(\mathscr{T}_R)]\right]\\ &= \sigma\left[\bigcup_{f\in C(X)} \sigma[f^{-1}(\mathscr{T}_R)]\right] \subset \sigma\left[\bigcup_{f\in C(X)} \mathscr{B}_X\right] = \mathscr{B}_X.\end{aligned}$$

To prove the reverse containment $\mathscr{B}_X \subset \mathscr{F}$ we need only establish that $\mathscr{F}$ contains every nonempty open set. By Lemma 7.2, it suffices to show that $\mathscr{F}$ contains every nonempty closed set. Let $A$ be such a set. We may assume without loss of generality that $A \neq X$, so there exists $x \in X - A$. Let $B = \{x\}$, and let $f$ be given by Lemma 7.1. Then $A = f^{-1}(\{0\}) \in \mathscr{F}$. Q.E.D.

We use Lemma 7.2 to prove another useful characterization of the Borel $\sigma$-algebra in a metrizable space.

**Proposition 7.11** Let $X$ be a metrizable space. Then $\mathscr{B}_X$ is the smallest class of sets which is closed under countable unions and intersections and contains every closed (open) set.

*Proof* Let $\mathscr{D}$ be the smallest class of sets which contains every closed set and is closed under countable unions and intersections, i.e., $\mathscr{D}$ is the intersection of all such classes. Then $\mathscr{D} \subset \mathscr{B}_X$ and it suffices to prove that $\mathscr{D}$ is closed under complementation. Let $\mathscr{D}'$ be the class of complements of sets in $\mathscr{D}$. Then $\mathscr{D}'$ is also closed under countable unions and intersections. Lemma 7.2 implies that $\mathscr{D}$ contains every open set, so $\mathscr{D}'$ contains every closed set, and consequently $\mathscr{D} \subset \mathscr{D}'$. Given $D \in \mathscr{D}$, we have $D \in \mathscr{D}'$, so $D^c \in \mathscr{D}$. Q.E.D.

**Definition 7.7** Let $X$ be a topological space. If there exists a complete separable metric space $Y$ and a Borel subset $B \in \mathscr{B}_Y$ such that $X$ is homeomorphic to $B$, then $X$ is said to be a *Borel space*. The empty set will also be regarded as a Borel space.

Note that every Borel space is metrizable and separable. Also, every complete separable metrizable space is a Borel space. Examples of Borel spaces are $R$, $R^n$, and $R^*$ with the weakest topology containing the intervals $[-\infty, \alpha)$, $(\beta, \infty]$, $(\alpha, \beta)$, $\alpha, \beta \in R$. (This is also the topology that makes the function $\varphi$ defined by

$$\varphi(x) = \begin{cases} 1 & \text{if } x = \infty,\\ \operatorname{sgn}(x)(1 - e^{-|x|}) & \text{if } x \in R,\\ -1 & \text{if } x = -\infty,\end{cases}$$

a homeomorphism from $R^*$ onto $[-1, 1]$). Any countable set $X$ with the discrete topology (i.e., the topology consisting of all subsets of $X$) is also a Borel space. We will show that every Borel subset of a Borel space is itself a Borel space. For this we shall need the following two lemmas. The proof of the first is elementary and is left to the reader.

**Lemma 7.4** If $Y$ is a topological space and $E \subset Y$, then the $\sigma$-algebra $\mathscr{B}_E$ generated by the relative topology coincides with the relative $\sigma$-algebra, i.e., the collection $\{E \cap C | C \in \mathscr{B}_Y\}$. In particular, if $E \in \mathscr{B}_Y$, then $\mathscr{B}_E$ consists of the Borel subsets of $Y$ contained in $E$.

**Lemma 7.5** If $X$ and $Y$ are topological spaces and $\varphi$ is a homeomorphism of $X$ into $Y$, then $\varphi(\mathscr{B}_X) = \mathscr{B}_{\varphi(X)}$.

*Proof* If $\mathscr{T}_X$ is the topology of $X$, then $\varphi(\mathscr{T}_X)$ is the topology of $\varphi(X)$. Since $\varphi$ is one-to-one, we have that $\varphi$ is the inverse of a mapping, and

$$\varphi(\mathscr{B}_X) = \varphi[\sigma(\mathscr{T}_X)] = \sigma[\varphi(\mathscr{T}_X)] = \mathscr{B}_{\varphi(X)}. \qquad \text{Q.E.D.}$$

**Proposition 7.12** If $X$ is a Borel space and $B \in \mathscr{B}_X$, then $B$ is a Borel space.

*Proof* Let $\varphi$ be a homeomorphism of $X$ into some complete separable metric space $Y$ such that $\varphi(X) \in \mathscr{B}_Y$. From Lemma 7.5 and the fact that $B \in \mathscr{B}_X$, we obtain $\varphi(B) \in \mathscr{B}_{\varphi(X)}$. It follows from Lemma 7.4 that $\varphi(B) \in \mathscr{B}_Y$. Q.E.D.

Like separability and completeness, the property of being a Borel space is preserved when taking countable Cartesian products.

**Proposition 7.13** Let $X_1, X_2, \ldots$ be a sequence of Borel spaces and $Y_n = X_1 X_2 \cdots X_n$, $Y = X_1 X_2 \cdots$. Then $Y$ and each $Y_n$ with the product topology is a Borel space and the Borel $\sigma$-algebras coincide with the product $\sigma$-algebras, i.e., $\mathscr{B}_{Y_n} = \mathscr{B}_{X_1}\mathscr{B}_{X_2} \cdots \mathscr{B}_{X_n}$ and $\mathscr{B}_Y = \mathscr{B}_{X_1}\mathscr{B}_{X_2} \cdots$.

*Proof* As in Proposition 7.4, we focus our attention on the more difficult infinite product. Consider the last statement of the proposition. Each $X_k$ has a countable base $\mathscr{G}_k$ for its topology, and the collection of sets of the form $G_1 G_2 \cdots G_n X_{n+1} X_{n+2} \cdots$, where $G_k$ ranges over $\mathscr{G}_k$ and $n$ ranges over the positive integers, is a base for the product topology on $Y$. The $\sigma$-algebra generated by this topology is $\mathscr{B}_Y$. Recall that the product $\sigma$-algebra $\mathscr{B}_{X_1}\mathscr{B}_{X_2} \cdots$ is the smallest $\sigma$-algebra containing all finite-dimensional measurable rectangles, i.e., all sets of the form $B_1 B_2 \cdots B_n X_{n+1}, X_{n+2} \cdots$, where $B_k \in \mathscr{B}_{X_k}$, $k = 1, \ldots, n$. It is clear that each basic set of the product topology on $Y$ is a finite-dimensional measurable rectangle, and since each open subset of $Y$ is a countable union of these basic open sets, every open subset of $Y$ is $\mathscr{B}_{X_1}\mathscr{B}_{X_2} \cdots$ measurable. We conclude that $\mathscr{B}_Y \subset \mathscr{B}_{X_1}\mathscr{B}_{X_2} \cdots$. (Note that this argument relies only on the separability of the spaces $X_1, X_2, \ldots$. Without this separability assumption, the argument fails and the conclusion is false.) The reverse set containment follows from the observation that for each $k$ and $B_k \in \mathscr{B}_{X_k}$, $X_1 X_2 \cdots X_{k-1} B_k X_{k+1} \cdots \in \mathscr{B}_Y$.

To prove that $Y$ is a Borel space, note that $X_k$ can be mapped by a homeomorphism $\varphi_k$ onto a Borel subset of a separable topologically complete space $\tilde{X}_k$. The product $\tilde{Y} = \tilde{X}_1 \tilde{X}_2 \cdots$ is separable and topologically complete, and $\varphi: Y \to \tilde{Y}$ defined by

$$\varphi(x_1, x_2, \ldots) = (\varphi_1(x_1), \varphi_2(x_2), \ldots)$$

is a homeomorphism from $Y$ onto $\varphi_1(X_1)\varphi_2(X_2) \cdots$. This last set is in $\mathscr{B}_{\tilde{X}_1} \mathscr{B}_{\tilde{X}_2} \cdots = \mathscr{B}_{\tilde{Y}}$, and the conclusion follows. Q.E.D.

**Definition 7.8** Let $X$ and $Y$ be topological spaces. A function $f: X \to Y$ is *Borel-measurable* if $f^{-1}(B) \in \mathscr{B}_X$ for every $B \in \mathscr{B}_Y$.

In many respects, Borel-measurable functions relate to Borel $\sigma$-algebras as continuous functions relate to topologies. We have already used the fact, for example, that if $f_k: X \to Y_k$ is continuous from a topological space $X$ to a topological space $Y_k$, $k = 1, 2, \ldots$, then $F: X \to Y_1 Y_2 \cdots$ defined by $F(x) = (f_1(x), f_2(x), \ldots)$ is also continuous. This follows from the componentwise nature of convergence in product spaces. There is an analogous fact for Borel-measurable functions and Borel spaces.

**Proposition 7.14** Let $X$ be a Borel space, $Y_1, Y_2, \ldots$ a sequence of Borel spaces, and $f_k: X \to Y_k$ a sequence of functions. If each $f_k$ is Borel-measurable, $k = 1, 2, \ldots$, then the function $F: X \to Y_1 Y_2 \cdots$ defined by

$$F(x) = (f_1(x), f_2(x), \ldots)$$

and the functions $F_n: X \to Y_1 Y_2 \cdots Y_n$ defined by

$$F_n(x) = (f_1(x), f_2(x), \ldots, f_n(x))$$

are Borel-measurable. Conversely, if $F$ is Borel-measurable, then each $f_k$ is Borel-measurable, $k = 1, 2, \ldots$, and if some $F_n$ is Borel-measurable, then $f_1, f_2, \ldots, f_n$ are Borel-measurable.

*Proof* Again we consider only the infinite product. The Borel $\sigma$-algebra in $Y_1 Y_2 \cdots$ is generated by sets of the form $B_1 B_2 \cdots$, where $B_k \in \mathscr{B}_{Y_k}$, $k = 1, 2, \ldots$. Now

$$F^{-1}(B_1 B_2 \cdots) = f_1^{-1}(B_1) \cap f_2^{-1}(B_2) \cap \cdots. \tag{19}$$

The left side of (19) is in $\mathscr{B}_X$ for each $B_k \in \mathscr{B}_{Y_k}$, $k = 1, 2, \ldots$, if and only if the sets $f_k^{-1}(B_k)$ are in $\mathscr{B}_X$ for each $B_k \in \mathscr{B}_{Y_k}$, $k = 1, 2, \ldots$, and the result follows. Q.E.D.

**Corollary 7.14.1** Let $X$ and $Y$ be Borel spaces, $D$ a Borel subset of $X$, and $f:D \to Y$ Borel-measurable. Then

$$\mathrm{Gr}(f) = \{(x, f(x)) \in XY | x \in D\}$$

is Borel-measurable.

*Proof* The mappings $(x, y) \to f(x)$ and $(x, y) \to y$ are Borel-measurable from $DY$ to $Y$, so the mapping $F(x, y) = (f(x), y)$ is Borel-measurable from $DY$ to $YY$. Then

$$\mathrm{Gr}(f) = F^{-1}(\{(y, y) | y \in Y\}).$$

Since $\{(y, y) | y \in Y\}$ is closed in $YY$, $\mathrm{Gr}(f)$ is Borel-measurable. Q.E.D.

The concept of homeomorphism is instrumental in classifying topological spaces, since it allows us to identify those which are "topologically equivalent." We can also classify measurable spaces by identifying those which, when regarded only as sets with $\sigma$-algebras, are indistinguishable. We specialize this concept to Borel spaces.

**Definition 7.9** Let $X$ and $Y$ be Borel spaces and $\varphi: X \to Y$ a Borel-measurable, one-to-one function such that $\varphi^{-1}$ is Borel-measurable on $\varphi(X)$. Then $\varphi$ is called a *Borel isomorphism*, and we say that $X$ and $\varphi(X)$ are *Borel-isomorphic* (or simply *isomorphic*).

If $X$ and $Y$ are Borel spaces and $\varphi: X \to Y$ is a Borel isomorphism, it is tempting to think of $X$ and $\varphi(X)$ as identical measurable spaces. The difficulty with this is that $X$ is a Borel space, but $\varphi(X)$ is not required to be. This discrepancy is eliminated by the following intuitively plausible proposition, the rather lengthy proof of which can be found in Chapter I, Section 3 of Parthasarathy [P1]. We will not have occasion to use this result.

**Proposition 7.15** (Kuratowski's theorem) Let $X$ be a Borel space, $Y$ a separable metrizable space, and $\varphi: X \to Y$ one-to-one and Borel-measurable. Then $\varphi(X)$ is a Borel subset of $Y$ and $\varphi^{-1}$ is Borel-measurable. In particular, if $Y$ is a Borel space, then $X$ and $\varphi(X)$ are isomorphic Borel spaces.

The advantage of classifying spaces by means of Borel isomorphisms is illustrated by the following result. We need this proposition for the subsequent development, but the proof is rather lengthy and is relegated to Appendix B, Section 2.

**Proposition 7.16** Let $X$ and $Y$ be Borel spaces. Then $X$ and $Y$ are isomorphic if and only if they have the same cardinality.

Proposition 7.16 leads to a consideration of the possible cardinalities of Borel spaces. Of course, Borel spaces which are countably infinite are possible, as are Borel spaces which consist of a given finite number of elements. In both these cases, the Borel $\sigma$-algebra is the power set and the conclusion of Proposition 7.16 is trivial. Because every Borel space can be homeomorphically embedded in the Hilbert cube, every Borel space has cardinality less than or equal to $c$. Even if one were to admit the possibility of an uncountable cardinality strictly less than $c$, the proof of Proposition 7.16 as given in Appendix B shows that every uncountable Borel space has cardinality $c$. By combining this fact with Proposition 7.16, we obtain the following corollary.

**Corollary 7.16.1** Every uncountable Borel space is Borel-isomorphic to every other uncountable Borel space. In particular, every uncountable Borel space is isomorphic to the unit interval $[0,1]$ and the Baire null space $\mathscr{N}$.

## 7.4 Probability Measures on Borel Spaces

If $X$ is a metrizable space, we shall refer to a probability measure $p$ on the measurable space $(X, \mathscr{B}_X)$ as simply a probability measure on $X$. The set of all probability measures on $X$ will be denoted by $P(X)$. A probability measure $p \in P(X)$ determines a linear functional $l_p : C(X) \to R$ defined by

$$l_p(f) = \int f dp. \tag{20}$$

On the other hand, a function $f \in C(X)$ determines a real-valued function $\theta_f : P(X) \to R$ defined by

$$\theta_f(p) = \int f\, dp. \tag{21}$$

These relationships and the metrizability of the underlying space $X$ allow us to show several properties of $P(X)$. In particular, we will prove that there is a natural topology on $P(X)$, the weakest topology with respect to which every mapping of the form of (21) is continuous, under which $P(X)$ is a Borel space whenever $X$ is a Borel space.

### *7.4.1 Characterization of Probability Measures*

**Definition 7.10** Let $X$ be a metrizable space. A probability measure $p \in P(X)$ is said to be *regular* if for every $B \in \mathscr{B}_X$,

$$p(B) = \sup\{p(F) | F \subset B,\ F \text{ closed}\} = \inf\{p(G) | B \subset G,\ G \text{ open}\}. \tag{22}$$

**Proposition 7.17** Let $X$ be a metrizable space. Every probability measure in $P(X)$ is regular.

*Proof* Let $p \in P(X)$ be given and let $\mathscr{E}$ be the collection of $B \in \mathscr{B}_X$ for which (22) holds. If $H \subset X$ is open, then $H = \bigcup_{n=1}^{\infty} F_n$, where $\{F_n\}$ is an increasing sequence of closed sets (Lemma 7.2), so

$$\begin{aligned}\inf\{p(G) \mid H \subset G,\ G \text{ open}\} &= p(H) \\ &= \lim_{n\to\infty} p(F_n) \\ &\le \sup\{p(F) \mid F \subset H,\ F \text{ closed}\} \le p(H).\end{aligned}$$

Therefore $\mathscr{E}$ contains every open subset of $X$. We show that $\mathscr{E}$ is a $\sigma$-algebra and conclude that $\mathscr{E} = \mathscr{B}_X$.

If $B \in \mathscr{E}$, then

$$\begin{aligned}p(B^c) = 1 - p(B) &= 1 - \sup\{p(F) \mid F \subset B,\ F \text{ closed}\} \\ &= \inf\{p(G) \mid B^c \subset G,\ G \text{ open}\},\end{aligned}$$

and similarly,

$$p(B^c) = \sup\{p(F) \mid F \subset B^c,\ F \text{ closed}\},$$

so $\mathscr{E}$ is closed under complementation. Now suppose $\{B_n\}$ is a sequence of sets in $\mathscr{E}$. Choose $\varepsilon > 0$ and $F_n \subset B_n \subset G_n$ such that $F_n$ is closed, $G_n$ is open, and $p(G_n - F_n) \le \varepsilon/2^n$. Then

$$\begin{aligned}\bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} G_n &= \left(\bigcup_{n=1}^{\infty} F_n\right) \cup \left[\bigcup_{n=1}^{\infty} (G_n - F_n)\right] \\ &\subset \left(\bigcup_{n=1}^{\infty} B_n\right) \cup \left[\bigcup_{n=1}^{\infty} (G_n - F_n)\right], \end{aligned} \tag{23}$$

so

$$p\left(\bigcup_{n=1}^{\infty} G_n\right) \le p\left(\bigcup_{n=1}^{\infty} B_n\right) + \varepsilon,$$

and since $\varepsilon$ is arbitrary,

$$p\left(\bigcup_{n=1}^{\infty} B_n\right) = \inf\left\{p(G) \,\middle|\, \bigcup_{n=1}^{\infty} B_n \subset G,\ G \text{ open}\right\}.$$

It is also apparent from (23) that

$$p\left(\bigcup_{n=1}^{\infty} B_n\right) \le p\left(\bigcup_{n=1}^{\infty} F_n\right) + \varepsilon,$$

so for $N$ sufficiently large,

$$p\left(\bigcup_{n=1}^{\infty} B_n\right) \le p\left(\bigcup_{n=1}^{N} F_n\right) + 2\varepsilon.$$

The finite union $\bigcup_{n=1}^{N} F_n$ is a closed subset of $\bigcup_{n=1}^{\infty} B_n$ and $\varepsilon$ is arbitrary, so

$$p\left(\bigcup_{n=1}^{\infty} B_n\right) = \sup\left\{p(F) \,\middle|\, F \subset \bigcup_{n=1}^{\infty} B_n,\ F \text{ closed}\right\}.$$

This shows that $\mathscr{E}$ is closed under countable unions and completes the proof. Q.E.D.

From Proposition 7.17 we conclude that a probability measure on a metrizable space is completely determined by its values on the open or closed sets. The following proposition is a similar result. It states that a probability measure $p$ on a metric space $(X, d)$ is completely determined by the values $\int g\, dp$, where $g$ ranges over $U_d(X)$.

**Proposition 7.18** Let $X$ be a metrizable space and $d$ a metric on $X$ consistent with its topology. If $p_1, p_2 \in P(X)$ and

$$\int g\, dp_1 = \int g\, dp_2 \qquad \forall g \in U_d(X),$$

then $p_1 = p_2$.

*Proof* Let $F$ be any closed proper subset of $X$ and let $G_n = \{x \in X | d(x, F) < 1/n\}$. For sufficiently large $n$, $F$ and $G_n^c$ are disjoint nonempty closed sets for which $\inf_{x \in F, y \in G_n^c} d(x, y) > 0$, so by Lemma 7.1, there exist functions $f_n \in U_d(X)$ such that $f_n(x) = 0$ for $x \in G_n^c$, $f_n(x) = 1$ for $x \in F$, and $0 \le f_n(x) \le 1$ for every $x \in X$. Then

$$p_1(F) \le \int f_n\, dp_1 = \int f_n\, dp_2 \le p_2(G_n),$$

and so

$$p_1(F) \le p_2\left(\bigcap_{n=1}^{\infty} G_n\right) = p_2(F).$$

Reversing the roles of $p_1$ and $p_2$, we obtain $p_1(F) = p_2(F)$. Proposition 7.17 implies $p_1(B) = p_2(B)$ for every $B \in \mathscr{B}_X$. Q.E.D.

### *7.4.2 The Weak Topology*

We turn now to a discussion of topologies on $P(X)$, where $X$ is a metrizable space. Given $\varepsilon > 0$, $p \in P(X)$, and $f \in C(X)$, define the subset of $P(X)$:

$$V_\varepsilon(p; f) = \left\{q \in P(X) \,\middle|\, \left|\int f\, dq - \int f\, dp\right| < \varepsilon\right\}. \tag{24}$$

If $D \subset C(X)$, consider the collection of subsets of $P(X)$:

$$\mathscr{V}(D) = \left\{ V_\varepsilon(p;f) \,\middle|\, \varepsilon > 0,\, p \in P(X),\, f \in D \right\}.$$

Let $\mathscr{T}(D)$ be the weakest topology on $P(X)$ which contains the collection $\mathscr{V}(D)$, i.e., the topology for which $\mathscr{V}(D)$ is a subbase.

**Lemma 7.6** Let $X$ be a metrizable space and $D \subset C(X)$. Let $\{p_\alpha\}$ be a net in $P(X)$ and $p \in P(X)$. Then $p_\alpha \to p$ relative to the topology $\mathscr{T}(D)$ if and only if $\int f\, dp_\alpha \to \int f\, dp$ for every $f \in D$.

*Proof* Suppose $p_\alpha \to p$ and $f \in D$. Then, given $\varepsilon > 0$, there exists $\beta$ such that $\alpha \geq \beta$ implies $p_\alpha \in V_\varepsilon(p;f)$. Hence $\int f\, dp_\alpha \to \int f\, dp$. Conversely, if $\int f\, dp_\alpha \to \int f\, dp$ for every $f \in D$, and $G \in \mathscr{T}(D)$ contains $p$, then $p$ is also contained in some basic open set $\bigcap_{k=1}^n V_{\varepsilon_k}(p;f_k) \subset G$, where $\varepsilon_k > 0$ and $f_k \in D$, $k = 1, \ldots, n$. Choose $\beta$ such that for all $\alpha \geq \beta$ we have $\left|\int f_k\, dp_\alpha - \int f_k\, dp\right| < \varepsilon_k$, $k = 1, \ldots, n$. Then $p_\alpha \in G$ for $\alpha \geq \beta$, so $p_\alpha \to p$. Q.E.D.

We are really interested in $\mathscr{T}[C(X)]$, the so-called *weak topology* on $P(X)$. The space $C(X)$ is too large to be manipulated easily, so we will need a countable set $D \subset C(X)$ such that $\mathscr{T}(D) = \mathscr{T}[C(X)]$. Such a set $D$ is produced by the next three lemmas.

**Lemma 7.7** Let $X$ be a metrizable space and $d$ a metric on $X$ consistent with its topology. If $f \in C(X)$, then there exist sequences $\{g_n\}$ and $\{h_n\}$ in $U_d(X)$ such that $g_n \uparrow f$ and $h_n \downarrow f$.

*Proof* We need only produce the sequence $\{g_n\}$, since the other case follows by considering $-f$. In Lemma 7.14 under weaker assumptions we will have occasion to utilize the construction about to be described, so we are careful to point out which assumptions are being used. If $f \in C(X)$, then $f$ is bounded below by some $b \in R$, and for at least one $x_0 \in X$, $f(x_0) < \infty$. Define

$$g_n(x) = \inf_{y \in X} [f(y) + nd(x,y)]. \tag{25}$$

Note that for every $x \in X$,

$$b \leq g_n(x) \leq f(x) + nd(x,x) = f(x),$$

and

$$b \leq g_n(x) \leq f(x_0) + nd(x,x_0) < \infty.$$

Thus

$$b \leq g_1 \leq g_2 \leq \cdots \leq f, \tag{26}$$

and each $g_n$ is finite-valued. For every $x, y, z \in X$,

$$f(y) + nd(x, y) \leq f(y) + nd(z, y) + nd(x, z),$$

and infimizing first the left side and then the right over $y \in X$, we obtain

$$g_n(x) \leq g_n(z) + nd(x, z).$$

Reverse the roles of $x$ and $z$ to show that

$$|g_n(x) - g_n(z)| \leq nd(x, z),$$

so $g_n \in U_d(X)$ for each $n$. From (26) we have

$$\lim_{n \to \infty} g_n \leq f. \tag{27}$$

We have so far used only the facts that $f$ is bounded below and not identically $\infty$. To prove that equality holds in (27), we use the continuity of $f$. For $x \in X$, and $\varepsilon > 0$, let $\{y_n\} \subset X$ be such that

$$f(y_n) + nd(x, y_n) \leq g_n(x) + \varepsilon.$$

As $n \to \infty$, either $g_n \uparrow \infty$, in which case equality must hold in (27), or else $y_n \to x$. In the latter case we have

$$f(x) = \lim_{n \to \infty} f(y_n) \leq \lim_{n \to \infty} g_n(x) + \varepsilon, \tag{28}$$

and since $x$ and $\varepsilon$ are arbitrary, equality holds in (27). Q.E.D.

**Lemma 7.8** Let $X$ be a metrizable space and $d$ a metric on $X$ consistent with its topology. Then $\mathscr{T}[C(X)] = \mathscr{T}[U_d(X)]$.

*Proof* Since $U_d(X) \subset C(X)$, we have $\mathscr{V}[U_d(X)] \subset \mathscr{V}[C(X)]$ and $\mathscr{T}[U_d(X)] \subset \mathscr{T}[C(X)]$. To prove the reverse containment, we show that every set in $\mathscr{V}[C(X)]$ is open in the $\mathscr{T}[U_d(X)]$ topology. Thus, given any set $V_\varepsilon(p; f) \in \mathscr{V}[C(X)]$ and any point $p_0$ in this set, we will construct a set in $\mathscr{T}[U_d(X)]$ containing $p_0$ and contained in $V_\varepsilon(p; f)$. Given $V_\varepsilon(p; f)$ and $p_0 \in V_\varepsilon(p; f)$, there exists $\varepsilon_0 > 0$ for which $V_{\varepsilon_0}(p_0; f) \subset V_\varepsilon(p; f)$. By Lemma 7.7, there exist functions $g$ and $h$ in $U_d(X)$ such that $g \leq f \leq h$ and

$$\int f \, dp_0 < \int g \, dp_0 + (\varepsilon_0/2), \qquad \int h \, dp_0 < \int f \, dp_0 + (\varepsilon_0/2).$$

If $q \in V_{\varepsilon_0/2}(p_0; g) \cap V_{\varepsilon_0/2}(p_0; h)$, then

$$\int f \, dp_0 < \int g \, dp_0 + (\varepsilon_0/2) < \int g \, dq + \varepsilon_0 \leq \int f \, dq + \varepsilon_0$$

and

$$\int f \, dq \leq \int h \, dq < \int h \, dp_0 + (\varepsilon_0/2) < \int f \, dp_0 + \varepsilon_0,$$

so

$$\left|\int f\,dq - \int f\,dp_0\right| < \varepsilon_0,$$

i.e., $q \in V_{\varepsilon_0}(p_0; f)$ and

$$V_{\varepsilon_0/2}(p_0; g) \cap V_{\varepsilon_0/2}(p_0; h) \subset V_\varepsilon(p; f). \qquad \text{Q.E.D.}$$

**Lemma 7.9** Let $X$ be a metrizable space and $d$ a metric on $X$ consistent with its topology. If $D$ is dense in $U_d(X)$, then $\mathscr{T}[U_d(X)] = \mathscr{T}(D)$.

*Proof* It is clear that $\mathscr{T}(D) \subset \mathscr{T}[U_d(X)]$. To prove the reverse set containment, we choose a set $V_\varepsilon(p; g) \in \mathscr{V}[U_d(X)]$, select a point $p_0$ in this set, and construct a set in $\mathscr{T}(D)$ containing $p_0$ and contained in $V_\varepsilon(p; g)$. Let

$$\varepsilon_0 = \varepsilon - \left|\int g\,dp_0 - \int g\,dp\right| > 0.$$

Let $h \in D$ be such that $\|g - h\| < \varepsilon_0/3$. Then for any $q \in V_{\varepsilon_0/3}(p_0; h)$, we have

$$\begin{aligned}\left|\int g\,dq - \int g\,dp\right| &\le \left|\int g\,dq - \int h\,dq\right| + \left|\int h\,dq - \int h\,dp_0\right| \\ &\quad + \left|\int h\,dp_0 - \int g\,dp_0\right| + \left|\int g\,dp_0 - \int g\,dp\right| \\ &< (\varepsilon_0/3) + (\varepsilon_0/3) + (\varepsilon_0/3) + \left|\int g\,dp_0 - \int g\,dp\right| = \varepsilon,\end{aligned}$$

so $V_{\varepsilon_0/3}(p_0; h) \subset V_\varepsilon(p; g)$. Q.E.D.

**Proposition 7.19** Let $X$ be a separable metrizable space. There is a metric $d$ on $X$ consistent with its topology and a countable dense subset $D$ of $U_d(X)$ such that $\mathscr{T}(D)$ is the weak topology $\mathscr{T}[C(X)]$ on $P(X)$.

*Proof* Corollary 7.6.1 states that the separable metrizable space $X$ has a totally bounded metrization $d$. By Proposition 7.9, there exists a countable dense set $D$ in $U_d(X)$. The conclusion follows from Lemmas 7.8 and 7.9. Q.E.D.

*From this point on, whenever $X$ is a metrizable space, we will understand $P(X)$ to be a topological space with the weak topology $\mathscr{T}[C(X)]$.* We will show that when $X$ is separable and metrizable, $P(X)$ is separable and metrizable; when $X$ is compact and metrizable, $P(X)$ is compact and metrizable; when $X$ is separable and topologically complete, $P(X)$ is separable and topologically complete; and when $X$ is a Borel space, $P(X)$ is a Borel space.

**Proposition 7.20** If $X$ is a separable metrizable space, then $P(X)$ is separable and metrizable.

*Proof* Let $d$ be a metric on $X$ consistent with its topology and $D$ a countable dense subset of $U_d(X)$ such that $\mathscr{T}(D)$ is the weak topology on $P(X)$ (Proposition 7.19). Let $R^\infty$ be the product of countably many copies of the real line and let $\varphi : P(X) \to R^\infty$ be defined by

$$\varphi(p) = \left( \int g_1 \, dp, \int g_2 \, dp, \ldots \right),$$

where $\{g_1, g_2, \ldots\}$ is an enumeration of $D$. We will show that $\varphi$ is a homeomorphism, and since $R^\infty$ is metrizable and separable (Proposition 7.4), these properties for $P(X)$ will follow.

Suppose that $\varphi(p_1) = \varphi(p_2)$, so that $\int g_k \, dp_1 = \int g_k \, dp_2$ for every $g_k \in D$. If $g \in U_d(X)$, then there exists a sequence $\{g_{k_j}\} \subset D$ such that $\|g_{k_j} - g\| \to 0$ as $j \to \infty$. Then

$$\begin{aligned} \left| \int g \, dp_1 - \int g \, dp_2 \right| &\le \limsup_{j \to \infty} \left| \int (g - g_{k_j}) \, dp_1 \right| + \limsup_{j \to \infty} \left| \int g_{k_j} \, dp_1 - \int g_{k_j} \, dp_2 \right| \\ &\quad + \limsup_{j \to \infty} \left| \int (g_{k_j} - g) \, dp_2 \right| \\ &\le 2 \limsup_{j \to \infty} \|g_{k_j} - g\| = 0, \end{aligned}$$

so $\int g \, dp_1 = \int g \, dp_2$. Proposition 7.18 implies that $p_1 = p_2$, so $\varphi$ is one-to-one. For each $g_k \in D$, the mapping $p \to \int g_k \, dp$ is continuous by Lemma 7.6, so $\varphi$ is continuous. To show that $\varphi^{-1}$ is continuous, let $\{p_\alpha\}$ be a net in $P(X)$ such that $\varphi(p_\alpha) \to \varphi(p)$ for some $p \in P(X)$. Then $\int g_k \, dp_\alpha \to \int g_k \, dp$ for every $g_k \in D$, and by Lemma 7.6, $p_\alpha \to p$. Q.E.D.

Proposition 7.20 guarantees that when $X$ is separable and metrizable, the topology on $P(X)$ can be characterized in terms of convergent sequences rather than nets. We give several conditions which are equivalent to convergence in $P(X)$.

**Proposition 7.21** Let $X$ be a separable metrizable space and let $d$ be a metric on $X$ consistent with its topology. Let $\{p_n\}$ be a sequence in $P(X)$ and $p \in P(X)$. The following statements are equivalent:

(a) $p_n \to p$;
(b) $\int f \, dp_n \to \int f \, dp$ for every $f \in C(X)$;
(c) $\int g \, dp_n \to \int g \, dp$ for every $g \in U_d(X)$;
(d) $\limsup_{n \to \infty} p_n(F) \le p(F)$ for every closed set $F \subset X$;
(e) $\liminf_{n \to \infty} p_n(G) \ge p(G)$ for every open set $G \subset X$.

*Proof* The equivalence of (a), (b), and (c) follows from Lemmas 7.6 and 7.8. The equivalence of (d) and (e) follows by complementation.

To show that (b) implies (d), let $F$ be a closed proper nonempty subset of $X$ and let $G_k = \{x \in X | d(x, F) < 1/k\}$. For $k$ sufficiently large, $F$ and $G_k^c$ are disjoint nonempty sets, and there exist functions $f_k \in C(X)$ such that $f_k(x) = 1$ for $x \in F$, $f_k(x) = 0$ for $x \in G_k^c$, and $0 \le f(x) \le 1$ for every $x \in X$. Using (b) we have

$$\limsup_{n\to\infty} p_n(F) \le \lim_{n\to\infty} \int f_k \, dp_n = \int f_k \, dp \le p(G_k),$$

and letting $k \to \infty$, we obtain (d).

To show that (d) implies (b), choose $f \in C(X)$ and assume without loss of generality that $0 \le f \le 1$. Choose a positive integer $K$ and define closed sets

$$F_k = \{x \in X | f(x) \ge k/K\}, \qquad k = 0, \ldots, K.$$

Define $\varphi : X \to [0, 1]$ by

$$\varphi(x) = \sum_{k=0}^{K} (k/K) \chi_{F_k - F_{k+1}}(x),$$

where $F_{K+1} = \varnothing$. Then $f - (1/K) \le \varphi \le f$, and, for any $q \in P(X)$,

$$\int \varphi \, dq = \sum_{k=0}^{K} (k/K) q(F_k - F_{k+1}) = (1/K) \sum_{k=1}^{K} q(F_k).$$

Using (d) we have

$$\begin{aligned}
\limsup_{n\to\infty} \int f \, dp_n - (1/K) &\le \limsup_{n\to\infty} \int \varphi \, dp_n \\
&= (1/K) \limsup_{n\to\infty} \sum_{k=1}^{K} p_n(F_k) \\
&\le (1/K) \sum_{k=1}^{K} p(F_k) = \int \varphi \, dp \le \int f \, dp,
\end{aligned}$$

and since $K$ is arbitrary, we obtain

$$\limsup_{n\to\infty} \int f \, dp_n \le \int f \, dp \tag{29}$$

for every $f \in C(X)$. In particular, (29) holds for $-f$, so

$$\liminf_{n\to\infty} \int f \, dp_n = -\limsup_{n\to\infty} \int (-f) \, dp_n \ge -\int (-f) \, dp = \int f \, dp. \tag{30}$$

Combine (29) and (30) to conclude (b). Q.E.D.

When $X$ is a metrizable space, we denote by $p_x$ the probability measure on $p(X)$ which assigns unit point mass to $x$, i.e., $p_x(B) = 1$ if and only if $x \in B$.

**Corollary 7.21.1** Let $X$ be a metrizable space. The mapping $\delta: X \to P(X)$ defined by $\delta(x) = p_x$ is a homeomorphism.

*Proof* It is clear that $\delta$ is one-to-one. Suppose $\{x_n\}$ is a sequence in $X$ and $x \in X$. If $x_n \to x$ and $G$ is an open subset of $X$, then there are two possibilities. Either $x \in G$, in which case $x_n \in G$ for sufficiently large $n$, so $\liminf_{n\to\infty} p_{x_n}(G) = 1 = p_x(G)$, or else $x \notin G$, in which case $\liminf_{n\to\infty} p_{x_n}(G) \geq 0 = p_x(G)$. Proposition 7.21 implies $p_{x_n} \to p_x$, so $\delta$ is continuous. On the other hand, if $p_{x_n} \to p_x$ and $G$ is an open neighborhood of $x$, then since $\liminf_{n\to\infty} p_{x_n}(G) \geq p_x(G) = 1$, we must have $x_n \in G$ for sufficiently large $n$, i.e., $x_n \to x$. This shows that $\delta$ is a homeomorphism. Q.E.D.

From Corollary 7.21.1 we see that $p_n$ can converge to $p$ in such a way that strict inequality holds in (d) and (e) of Proposition 7.21. For example, let $G \subset X$ be open, let $x$ be on the boundary of $G$, and let $x_n$ converge to $x$ through $G$. Then $p_{x_n}(G) = 1$ for every $n$, but $p_x(G) = 0$.

We now show that compactness of $X$ is inherited by $P(X)$.

**Proposition 7.22** If $X$ is a compact metrizable space, then $P(X)$ is a compact metrizable space.

*Proof* If $X$ is a compact metrizable space, it is separable (Corollary 7.6.2) and $C(X)$ is separable (Proposition 7.7). Let $\{f_k\}$ be a countable set in $C(X)$ such that $f_1 \equiv 1$, $\|f_k\| \leq 1$ for every $k$, and $\{f_k\}$ is dense in the unit sphere $\{f \in C(X) \mid \|f\| \leq 1\}$. Let $[-1,1]^\infty$ be the product of countably many copies of $[-1,1]$ and define $\varphi: P(X) \to [-1,1]^\infty$ by

$$\varphi(p) = \left(\int f_1\,dp, \int f_2\,dp, \ldots\right).$$

A trivial modification of the proof of Proposition 7.20 shows $\varphi$ is a homeomorphism. We will show that $\varphi[P(X)]$ is closed in the compact space $[-1,1]^\infty$, and the compactness of $P(X)$ will follow.

Suppose $\{p_n\}$ is a sequence in $P(X)$ and $\varphi(p_n) \to (\alpha_1, \alpha_2, \ldots) \in [-1,1]^\infty$. Given $\varepsilon > 0$ and $f \in C(X)$ with $\|f\| \leq 1$, there is a function $f_k$ with $\|f - f_k\| < \varepsilon/3$. There is a positive integer $N$ such that $n, m \geq N$ implies $|\int f_k\,dp_n - \int f_k\,dp_m| < \varepsilon/3$. Then

$$\left|\int f\,dp_n - \int f\,dp_m\right| \leq \left|\int f\,dp_n - \int f_k\,dp_n\right| + \left|\int f_k\,dp_n - \int f_k\,dp_m\right| + \left|\int f_k\,dp_m - \int f\,dp_m\right| < \varepsilon,$$

so $\{\int f\,dp_n\}$ is Cauchy in $[-1,1]$. Denote its limit by $E(f)$. If $\|f\| > 1$, define

$$E(f) = \|f\|E(f/\|f\|).$$

It is easily verified that $E$ is a linear functional on $C(X)$, that $E(f) \geq 0$ whenever $f \geq 0$, $|E(f)| \leq \|f\|$ for every $f \in C(X)$, and $E(f_1) = 1$. Suppose $\{h_n\}$ is a sequence in $C(X)$ and $h_n(x) \downarrow 0$ for every $x \in X$. Then for each $\varepsilon > 0$, the set $K_n(\varepsilon) = \{x | h_n(x) \geq \varepsilon\}$ is compact, and $\bigcap_{n=1}^{\infty} K_n(\varepsilon) = \varnothing$. Therefore, for $n$ sufficiently large, $K_n(\varepsilon) = \varnothing$, which implies $\|h_n\| \downarrow 0$. Consequently, $E(h_n) \downarrow 0$. This shows that the functional $E$ is a *Daniell integral*, and by a classical theorem (see, e.g., Royden [R5, p. 299, Proposition 21]) there exists a unique probability measure on $\sigma[\bigcup_{f \in C(X)} f^{-1}(\mathscr{B}_R)]$ which satisfies $E(f) = \int f\,dp$ for every $f \in C(X)$. Proposition 7.10 implies $p \in P(X)$. We have

$$\alpha_k = \lim_{n\to\infty} \int f_k\,dp_n = E(f_k) = \int f_k\,dp, \qquad k = 1, 2, \ldots,$$

so $\varphi(p_n) \to \varphi(p)$. This proves $\varphi[P(X)]$ is closed. Q.E.D.

In order to show that toplogical completeness and separability of $X$ imply the same properties for $P(X)$, we need the following lemma.

**Lemma 7.10** Let $X$ and $Y$ be separable metrizable spaces and $\varphi: X \to Y$ a homeomorphism. The mapping $\psi: P(X) \to P(Y)$ defined by

$$\psi(p)(B) = p[\varphi^{-1}(B)] \qquad \forall B \in \mathscr{B}_Y$$

is a homeomorphism.

*Proof* Suppose $p_1, p_2 \in P(X)$ and $p_1 \neq p_2$. Since $p_1$ and $p_2$ are regular, there is an open set $G \subset X$ for which $p_1(G) \neq p_2(G)$. The image $\varphi(G)$ is relatively open in $\varphi(X)$, so $\varphi(G) = \varphi(X) \cap B$, where $B$ is open in $Y$. It is clear that

$$\psi(p_1)(B) = p_1(G) \neq p_2(G) = \psi(p_2)(B),$$

so $\psi$ is one-to-one. Let $\{p_n\}$ be a sequence in $P(X)$ and $p \in P(X)$. If $p_n \to p$, then since $\varphi^{-1}(H)$ is open in $X$ for every open set $H \subset Y$, Proposition 7.21 implies

$$\liminf_{n\to\infty} \psi(p_n)(H) = \liminf_{n\to\infty} p_n[\varphi^{-1}(H)] \geq p[\varphi^{-1}(H)] = \psi(p)(H),$$

so $\psi(p_n) \to \psi(p)$ and $\psi$ is continuous. If we are given $\{p_n\}$ and $p$ such that $\psi(p_n) \to \psi(p)$, a reversal of this argument shows that $p_n \to p$ and $\psi^{-1}$ is continuous. Q.E.D.

**Proposition 7.23** If $X$ is a topologically complete separable space, then $P(X)$ is topologically complete and separable.

*Proof* By Urysohn's theorem (Proposition 7.2) there is a homeomorphism $\varphi: X \to \mathscr{H}$, and the mapping $\psi$ obtained by replacing $Y$ by $\mathscr{H}$ in Lemma 7.10 is a homeomorphism from $P(X)$ to $P(\mathscr{H})$. Alexandroff's theorem (Proposition 7.3) implies $\varphi(X)$ is a $G_\delta$-subset of $\mathscr{H}$, and we see that

$$\psi[P(X)] = \{p \in P(\mathscr{H}) | p[\mathscr{H} - \varphi(X)] = 0\}. \tag{31}$$

We will show $\psi[P(X)]$ is a $G_\delta$-subset of the compact space $P(\mathscr{H})$ (Proposition 7.22) and use Alexandroff's theorem again to conclude that $P(X)$ is topologically complete.

Since $\varphi(X)$ is a $G_\delta$-subset of $\mathscr{H}$, we can find open sets $G_1 \supset G_2 \supset \cdots$ such that $\varphi(X) = \bigcap_{n=1}^{\infty} G_n$. It is clear from (31) that

$$\begin{aligned} \psi[P(X)] &= \bigcap_{n=1}^{\infty} \{p \in P(\mathscr{H}) | p(\mathscr{H} - G_n) = 0\} \\ &= \bigcap_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \{p \in P(\mathscr{H}) | p(\mathscr{H} - G_n) < 1/k\}. \end{aligned}$$

But for any closed set $F$ and real number $c$, the set $\{p \in P(\mathscr{H}) | p(F) \geq c\}$ is closed by Proposition 7.21(d), and $\{p \in P(\mathscr{H}) | p(\mathscr{H} - G_n) < 1/k\}$ is the complement of such a set. Q.E.D.

We turn now to characterizing the $\sigma$-algebra $\mathscr{B}_{P(X)}$ when $X$ is metrizable and separable. From Lemma 7.6, we have that the mapping $\theta_f: P(X) \to R$ given by

$$\theta_f(p) = \int f \, dp$$

is continuous for every $f \in C(X)$. One can easily verify from Proposition 7.21 that the mapping $\theta_B: P(X) \to [0, 1]$ defined by†

$$\theta_B(p) = p(B)$$

is Borel-measurable when $B$ is a closed subset of $X$. (Indeed, in the final stage of the proof of Proposition 7.23, we used the fact that when $B$ is closed the upper level sets $\{p \in P(X) | \theta_B(p) \geq c\}$ are closed.) Likewise, when $B$ is open, $\theta_B$ is Borel-measurable. It is natural to ask if $\theta_B$ is also Borel-measurable when $B$ is an arbitrary Borel set. The answer to this is yes, and in fact, $\mathscr{B}_{P(X)}$ is the smallest $\sigma$-algebra with respect to which $\theta_B$ is measurable for every $B \in \mathscr{B}_X$. A useful aid in proving this and several subsequent results is the concept of a Dynkin system.

† The use of the symbol $\theta_B$ here is a slight abuse of notation. In keeping with the definition of $\theta_f$, the technically correct symbol would be $\theta_{\chi_B}$.

**Definition 7.11** Let $X$ be a set and $\mathscr{D}$ a class of subsets of $X$. We say $\mathscr{D}$ is a *Dynkin system* if the following conditions hold:

(a) $X \in \mathscr{D}$.
(b) If $A, B \in \mathscr{D}$ and $B \subset A$, then $A - B \in \mathscr{D}$.
(c) If $A_1, A_2, \ldots \in \mathscr{D}$ and $A_1 \subset A_2 \subset \cdots$, then $\bigcup_{n=1}^{\infty} A_n \in \mathscr{D}$.

**Proposition 7.24** (Dynkin system theorem) Let $\mathscr{F}$ be a class of subsets of a set $X$, and assume $\mathscr{F}$ is closed under finite intersections. If $\mathscr{D}$ is a Dynkin system containing $\mathscr{F}$, then $\mathscr{D}$ also contains $\sigma(\mathscr{F})$.

*Proof* This is a standard result in measure theory. See, for example, Ash [A1, page 169]. Q.E.D.

**Proposition 7.25** Let $X$ be a separable metrizable space and $\mathscr{E}$ a collection of subsets of $X$ which generates $\mathscr{B}_X$ and is closed under finite intersections. Then $\mathscr{B}_{P(X)}$ is the smallest $\sigma$-algebra with respect to which all functions of the form

$$\theta_E(p) = p(E), \qquad E \in \mathscr{E},$$

are measurable from $P(X)$ to $[0, 1]$, i.e.,

$$\mathscr{B}_{P(X)} = \sigma\left[\bigcup_{E \in \mathscr{E}} \theta_E^{-1}(\mathscr{B}_R)\right].$$

*Proof* Let $\mathscr{F}$ be the smallest $\sigma$-algebra with respect to which $\theta_E$ is measurable for every $E \in \mathscr{E}$. To show $\mathscr{F} \subset \mathscr{B}_{P(X)}$, we show that $\theta_B$ is $\mathscr{B}_{P(X)}$-measurable for every $B \in \mathscr{B}_X$. Let $\mathscr{D} = \{B \in \mathscr{B}_X | \theta_B \text{ is } \mathscr{B}_{P(X)}\text{-measurable}\}$. It is easily verified that $\mathscr{D}$ is a Dynkin system. We have already seen that $\mathscr{D}$ contains every closed set, so the Dynkin system theorem (Proposition 7.24) implies $\mathscr{D} = \mathscr{B}_X$.

It remains to show that $\mathscr{B}_{P(X)} \subset \mathscr{F}$. Let $\mathscr{D}' = \{B \in \mathscr{B}_X | \theta_B \text{ is } \mathscr{F}\text{-measurable}\}$ As before, $\mathscr{D}'$ is a Dynkin system, and since $\mathscr{E} \subset \mathscr{D}'$, we have $\mathscr{D}' = \mathscr{B}_X$. Thus the function $\theta_f(p) = \int f\, dp$ is $\mathscr{F}$-measurable when $f$ is the indicator of a Borel set. Therefore $\theta_f$ is $\mathscr{F}$-measurable when $f$ is a Borel-measurable simple function. If $f \in C(X)$, then there is a sequence of simple functions $f_n$ which are uniformly bounded below such that $f_n \uparrow f$. The monotone convergence theorem implies $\theta_{f_n} \uparrow \theta_f$, so $\theta_f$ is $\mathscr{F}$-measurable. It follows that for $\varepsilon > 0$, $p \in P(X)$, and $f \in C(X)$, the subbasic open set

$$V_\varepsilon(p; f) = \left\{q \in P(X) \middle| \left|\int f\, dq - \int f\, dp\right| < \varepsilon\right\}$$

is $\mathscr{F}$-measurable. It follows that $\mathscr{B}_{P(X)} = \mathscr{F}$ (see the remark following Definition 7.6). Q.E.D.

**Corollary 7.25.1** If $X$ is a Borel space, then $P(X)$ is a Borel space.

*Proof* Let $\varphi$ be a homeomorphism mapping $X$ onto a Borel subset of a topologically complete separable space $Y$. Then, by Lemma 7.10, $P(X)$ is homeomorphic to the Borel set $\{p \in P(Y) \mid p[\varphi(X)] = 1\}$. Since $P(Y)$ is topologically complete and separable (Proposition 7.23), the result follows. Q.E.D.

*7.4.3 Stochastic Kernels*

We now consider probability measures on a separable metrizable space parameterized by the elements of another separable metrizable space.

**Definition 7.12** Let $X$ and $Y$ be separable metrizable spaces. A *stochastic kernel* $q(dy|x)$ on $Y$ given $X$ is a collection of probability measures in $P(Y)$ parameterized by $x \in X$. If $\mathscr{F}$ is a $\sigma$-algebra on $X$ and $\gamma^{-1}[\mathscr{B}_{P(Y)}] \subset \mathscr{F}$, where $\gamma: X \to P(Y)$ is defined by

$$\gamma(x) = q(dy|x), \tag{32}$$

then $q(dy|x)$ is said to be *$\mathscr{F}$-measurable.* If $\gamma$ is continuous, $q(dy|x)$ is said to be *continuous.*

**Proposition 7.26** Let $X$ and $Y$ be Borel spaces, $\mathscr{E}$ a collection of subsets of $Y$ which generates $\mathscr{B}_Y$ and is closed under finite intersections, and $q(dy|x)$ a stochastic kernel on $Y$ given $X$. Then $q(dy|x)$ is Borel-measurable if and only if the mapping $\lambda_E: X \to [0, 1]$ defined by

$$\lambda_E(x) = q(E|x)$$

is Borel-measurable for every $E \in \mathscr{E}$.

*Proof* Let $\gamma: X \to P(Y)$ be defined by $\gamma(x) = q(dy|x)$. Then for $E \in \mathscr{E}$, we have $\lambda_E = \theta_E \circ \gamma$. If $q(dy|x)$ is Borel-measurable (i.e., $\gamma$ is Borel-measurable), then Proposition 7.25 implies $\lambda_E$ is Borel-measurable for every $E \in \mathscr{E}$. Conversely, if $\lambda_E$ is Borel-measurable for every $E \in \mathscr{E}$, then $\sigma[\bigcup_{E \in \mathscr{E}} \lambda_E^{-1}(\mathscr{B}_R)] \subset \mathscr{B}_X$. Proposition 7.25 implies

$$\begin{aligned}\gamma^{-1}[\mathscr{B}_{P(Y)}] &= \gamma^{-1}\left[\sigma\left(\bigcup_{E \in \mathscr{E}} \theta_E^{-1}(\mathscr{B}_R)\right)\right] \\ &= \sigma\left[\bigcup_{E \in \mathscr{E}} \gamma^{-1}(\theta_E^{-1}(\mathscr{B}_R))\right] = \sigma\left[\bigcup_{E \in \mathscr{E}} \lambda_E^{-1}(\mathscr{B}_R)\right] \subset \mathscr{B}_X,\end{aligned}$$

so $q(dy|x)$ is Borel-measurable. Q.E.D.

**Corollary 7.26.1** Let $X$ and $Y$ be Borel spaces and $q(dy|x)$ a Borel-measurable stochastic kernel on $Y$ given $X$. If $B \in \mathscr{B}_{XY}$, then the mapping

$\Lambda_B: X \to [0,1]$ defined by

$$\Lambda_B(x) = q(B_x|x), \tag{33}$$

where $B_x = \{y \in Y | (x, y) \in B\}$, is Borel-measurable.

*Proof* If $B \in \mathscr{B}_{XY}$ and $x \in X$, then $B_x \subset Y$ is homeomorphic to $B \cap [\{x\}Y] \in \mathscr{B}_{XY}$. It follows that $B_x \in \mathscr{B}_Y$, so $q(B_x|x)$ is defined. It is easy to show that the collection $\mathscr{D} = \{B \in \mathscr{B}_{XY} | \Lambda_B$ is Borel-measurable$\}$ is a Dynkin system. Proposition 7.26 implies that $\mathscr{D}$ contains the measurable rectangles, so $\mathscr{D} = \mathscr{B}_{XY}$. Q.E.D.

We now show that one can decompose a probability measure on a product of Borel spaces into a marginal and a Borel-measurable stochastic kernel. This decomposition is possible even when a measurable dependence on a parameter is admitted, and, as we shall see in Chapter 10, this result is essential to the filtering algorithm for imperfect state information dynamic programming models.

As a notational convenience, we use $\underline{X}$ to denote a typical Borel subset of a Borel space $X$.

**Proposition 7.27** Let $(X, \mathscr{F})$ be a measurable space, let $Y$ and $Z$ be Borel spaces, and let $q(d(y,z)|x)$ be a stochastic kernel on $YZ$ given $X$. Assume that $q(B|x)$ is $\mathscr{F}$-measurable in $x$ for every $B \in \mathscr{B}_{YZ}$. Then there exists a stochastic kernel $r(dz|x,y)$ on $Z$ given $XY$ and a stochastic kernel $s(dy|x)$ on $Y$ given $X$ such that $r(\underline{Z}|x,y)$ is $\mathscr{F}\mathscr{B}_Y$-measurable in $(x,y)$ for every $\underline{Z} \in \mathscr{B}_Z$, $s(\underline{Y}|x)$ is $\mathscr{F}$-measurable in $x$ for every $\underline{Y} \in \mathscr{B}_Y$, and

$$q(\underline{Y}\underline{Z}|x) = \int_{\underline{Y}} r(\underline{Z}|x,y)s(dy|x) \qquad \forall \underline{Y} \in \mathscr{B}_Y, \quad \underline{Z} \in \mathscr{B}_Z. \tag{34}$$

*Proof* We prove this proposition under the assumption that $Y$ and $Z$ are uncountable. If either $Y$ or $Z$ or both are countable, slight modifications (actually simplifications) of this proof are necessary. From Corollary 7.16.1, we may assume without loss of generality that $Y = Z = (0,1]$.

Let $s(dy|x)$ be the marginal of $q(d(y,z)|x)$ on $Y$, i.e., $s(\underline{Y}|x) = q(\underline{Y}Z|x)$ for every $\underline{Y} \in \mathscr{B}_Y$. For each positive integer $n$, define subsets of $Y$

$$M(j,n) = ((j-1)/2^n, j/2^n], \qquad j = 1, \ldots, 2^n.$$

Then each $M(j, n+1)$ is a subset of some $M(k,n)$, and the collection $\{M(j,n) | n = 1, 2, \ldots; j = 1, \ldots, 2^n\}$ generates $\mathscr{B}_Y$. For $z \in Q \cap Z$, define $q(dy(0,z]|x)$ to be the measure on $Y$ whose value at $\underline{Y} \in \mathscr{B}_Y$ is $q(\underline{Y}(0,z]|x)$. Then $q(dy(0,z]|x)$ is absolutely continuous with respect to $s(dy|x)$ for every

$z \in Q \cap Z$ and $x \in X$. Define for $z \in Q \cap Z$

$$G_n(z|x,y) = \begin{cases} q[M(j,n)(0,z]|x]/s[M(j,n)|x] \\ \qquad \text{if} \quad y \in M(j,n) \quad \text{and} \quad s[M(j,n)|x] > 0, \\ 0 \qquad \text{if} \quad y \in M(j,n) \quad \text{and} \quad s[M(j,n)|x] = 0. \end{cases}$$

The functions $G_n(z|x,y)$ can be regarded as generalized difference quotients of $q(dy(0,z]|x)$ relative to $s(dy|x)$. For each $z$, the set

$$\begin{aligned} B(z) &= \left\{(x,y) \in XY \,\middle|\, \lim_{n\to\infty} G_n(z|x,y) \text{ exists in } R\right\} \\ &= \{(x,y) \in XY | \{G_n(z|x,y)\} \text{ is Cauchy}\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n,m\geq N} \{(x,y)\in XY \,|\, |G_n(z|x,y) - G_m(z|x,y)| < 1/k\} \end{aligned}$$

is $\mathscr{F}\mathscr{B}_Y$-measurable. Theorem 2.5, page 612 of Doob [D4] states that

$$s[B(z)_x|x] = 1 \qquad \forall x \in X, \quad z \in Q \cap Z,$$

and if we define

$$G(z|x,y) = \begin{cases} \lim_{n\to\infty} G_n(z|x,y) & \text{if} \quad (x,y) \in B(z), \\ z & \text{otherwise,} \end{cases}$$

then

$$q[\underline{Y}(0,z]|x] = \int_{\underline{Y}} G(z|x,y)s(dy|x) \qquad \forall x \in X, \quad z \in Q \cap Z, \quad \underline{Y} \in \mathscr{B}_Y. \quad (35)$$

It is clear that for any $z$, $G(z|x,y)$ is $\mathscr{F}\mathscr{B}_Y$-measurable in $(x,y)$.[†]

A comparison of (34) and (35) suggests that we should try to extend $G(z|x,y)$ in such a way that for fixed $(x,y)$, $G(z|x,y)$ is a distribution function.

[†] For the reader familiar with martingales, we give the proof of the theorem just referenced. Fix $x$ and $y$ and observe that for $m \geq n$,

$$q[M(j,n)(0,z]|x] = \int_{M(j,n)} G_m(z|x,y)s(dy|x). \qquad (*)$$

Since $\{M(j,n)|j = 1,\ldots,2^n\}$ is the $\sigma$-algebra generated by $G_n(z|x,y)$ regarded as a function of $y$, we conclude that $G_n(z|x,y)$, $n = 1,2,\ldots$ is a martingale on $Y$ under the measure $s(dy|x)$. Each $G_n(z|x,y)$ is bounded above by 1, so by the martingale convergence theorem (see, e.g., Ash [A1, p. 292]), $G_n(z|x,y)$ converges for $s(dy|x)$ almost every $y$. Thus $s[B(z)_x|x] = 1$ and the definition of $G(z|x,y)$ given above is possible. Let $m \to \infty$ in $(*)$ to see that (35) holds whenever $\underline{Y} = M(j,n)$ for some $j$ and $n$. The collection of sets $\underline{Y}$ for which (35) holds is a Dynkin system, and it follows from Proposition 7.24 that (35) holds for every $\underline{Y} \in \mathscr{B}_Y$.

Toward this end, for each $z_0 \in Q \cap Z$, we define

$$C(z_0) = \{(x, y) \in XY | \exists z \in Q \cap Z \text{ with } z \leq z_0 \text{ and } G(z|x, y) > G(z_0|x, y)\},$$
$$= \bigcup_{\substack{z \in Q \cap Z \\ z \leq z_0}} \{(x, y) \in XY | G(z|x, y) > G(z_0|x, y)\},$$
$$C = \bigcup_{z_0 \in Q \cap Z} C(z_0),$$
$$D(z_0) = \{(x, y) \in XY | G(\cdot|x, y) \text{ is not right-continuous at } z_0\}$$
$$= \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{\substack{z \in Q \cap Z \\ z_0 \leq z < z_0 + 1/k}} \{(x, y) \in XY \,|\, |G(z|x, y) - G(z_0|x, y)| \geq 1/n\},$$
$$D = \bigcup_{z_0 \in Q \cap Z} D(z_0),$$
$$E = \{(x, y) \in XY | G(z|x, y) \text{ does not converge to zero as } z \downarrow 0\}$$
$$= \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{\substack{z \in Q \cap Z \\ z < 1/k}} \{(x, y) \in XY \,|\, |G(z|x, y)| \geq 1/n\},$$

and

$$F = \{(x, y) \in XY | G(1|x, y) \neq 1\}.$$

For fixed $x \in X$ and $z_0 \in Q \cap Z$, (35) implies that whenever $z \in Q \cap Z$, $z \leq z_0$, then

$$\int_{\underline{Y}} G(z|x, y)s(dy|x) \leq \int_{\underline{Y}} G(z_0|x, y)s(dy|x) \qquad \forall \underline{Y} \in \mathscr{B}_Y.$$

Therefore $G(z|x, y) \leq G(z_0|x, y)$ for $s(dy|x)$ almost all $y$, so $s[C(z_0)_x|x] = 0$ and

$$s(C_x|x) = 0. \tag{36}$$

Equation (36) implies that $G(z|x, y)$ is nondecreasing in $z$ for $s(dy|x)$ almost all $y$. This fact and (35) imply that if $z \downarrow z_0$ ($z \in Q \cap Z$), then

$$\int_Y G(z|x, y)s(dy|x) \downarrow \int_Y G(z_0|x, y)s(dy|x),$$

and

$$G(z|x, y) \downarrow G(z_0|x, y)$$

for $s(dy|x)$ almost all $y$. Therefore $s[D(z_0)_x|x] = 0$ and

$$s(D_x|x) = 0. \tag{37}$$

Equation (35) also implies that as $z \downarrow 0$ $(z \in Q \cap Z)$

$$\int_{\underline{Y}} G(z|x,y)s(dy|x) \downarrow 0 \qquad \forall \underline{Y} \in \mathscr{B}_Y.$$

Since $G(z|x,y)$ is nondecreasing in $z$ for $s(dy|x)$ almost all $y$, we must have $G(z|x,y) \downarrow 0$ for $s(dy|x)$ almost all $y$, i.e.,

$$s(E_x|x) = 0. \tag{38}$$

Substituting $z = 1$ in (35), we see that

$$\int_{\underline{Y}} G(1|x,y)s(dy|x) = s(\underline{Y}|x) \qquad \forall \underline{Y} \in \mathscr{B}_Y,$$

so $G(1|x,y) = 1$ for $s(dy|x)$ almost all $y$, i.e.,

$$s(F_x|x) = 0. \tag{39}$$

For $z \in Z$, let $\{z_n\}$ be a sequence in $Q \cap Z$ such that $z_n \downarrow z$ and define, for every $x \in X$, $y \in Y$,

$$F(z|x,y) = \begin{cases} \lim_{n\to\infty} G(z_n|x,y) & \text{if} \quad (x,y) \in XY - (C \cup D \cup E \cup F), \\ z & \text{otherwise.} \end{cases} \tag{40}$$

For $(x,y) \in XY - (C \cup D \cup E \cup F)$, $G(z|x,y)$ is a nondecreasing right-continuous function of $z \in Q \cap Z$, so $F(z|x,y)$ is well defined, nondecreasing, and right-continuous. It also satisfies for every $(x,y) \in XY$,

$$0 \le F(z|x,y) \le 1 \qquad \forall z \in Z,$$
$$F(1|x,y) = 1,$$

and

$$\lim_{z \downarrow 0} F(z|x,y) = 0.$$

It is a standard result of probability theory (Ash [A1, p. 24]) that for each $(x,y)$ there is a probability measure $r(dz|x,y)$ on $Z$ such that

$$r((0,z]|x,y) = F(z|x,y) \qquad \forall z \in (0,1].$$

The collection of subsets $\underline{Z} \in \mathscr{B}_Z$ for which $r(\underline{Z}|x,y)$ is $\mathscr{F}\mathscr{B}_Y$-measurable in $(x,y)$ forms a Dynkin system which contains $\{(0,z]|z \in Z\}$, so $r(\underline{Z}|x,y)$ is $\mathscr{F}\mathscr{B}_Y$-measurable for every $\underline{Z} \in \mathscr{B}_Z$. Relations (35)–(40) and the monotone convergence theorem imply

$$\begin{aligned} q[\underline{Y}(0,z]|x] &= \int_{\underline{Y}} F(z|x,y)s(dy|x) \\ &= \int_{\underline{Y}} r((0,z]|x,y)s(dy|x) \qquad \forall x \in X, \quad z \in Z, \quad \underline{Y} \in \mathscr{B}_Y. \end{aligned} \tag{41}$$

The collection of subsets $\underline{Z} \in \mathscr{B}_Z$ for which (34) holds forms a Dynkin system which contains $\{(0, z] | z \in Z\}$, so (34) holds for every $\underline{Z} \in \mathscr{B}_Z$. Q.E.D.

If $\mathscr{F} = \mathscr{B}_X$, an application of Proposition 7.26 reduces Proposition 7.27 to the following form.

**Corollary 7.27.1** Let $X$, $Y$, and $Z$ be Borel spaces and let $q(d(y,z)|x)$ be a Borel-measurable stochastic kernel on $YZ$ given $X$. Then there exist Borel-measurable stochastic kernels $r(dz|x, y)$ and $s(dy|x)$ on $Z$ given $XY$ and on $Y$ given $X$, respectively, such that (34) holds.

If there is no dependence on the parameter $x$ in Corollary 7.27.1, we have the following well-known result for Borel spaces.

**Corollary 7.27.2** Let $Y$ and $Z$ be Borel spaces and $q \in P(YZ)$. Then there exists a Borel-measurable stochastic kernel $r(dz|y)$ on $Z$ given $Y$ such that

$$q(\underline{Y}\underline{Z}) = \int_{\underline{Y}} r(\underline{Z}|y)s(dy) \qquad \forall \underline{Y} \in \mathscr{B}_Y, \quad \underline{Z} \in \mathscr{B}_Z,$$

where $s$ is the marginal of $q$ on $Y$.

*7.4.4 Integration*

As in Section 2.1, we adopt the convention

$$-\infty + \infty = +\infty - \infty = \infty. \tag{42}$$

With this convention, for $a$, $b$, $c \in R^*$ the associative law

$$(a + b) + c = a + (b + c)$$

still holds, since if either $a$, $b$, or $c$ is $\infty$, then both sides of (42) are $\infty$, while if neither $a$, $b$, nor $c$ is $\infty$, the usual arithmetic involving finite numbers and $-\infty$ applies. Also, if $a, b, c \in R^*$ and $a + b = c$, then $a = c - b$, provided $b \neq \pm\infty$. It is always true however that if $a + b \leq c$, then $a \leq c - b$.

We use convention (42) to extend the definition of the integral. If $X$ is a metrizable space, $p \in P(X)$, and $f: X \to R^*$ is Borel-measurable, we define

$$\int f\,dp = \int f^+\,dp - \int f^-\,dp. \tag{43}$$

Note that if $\int f^+\,dp < \infty$ or if $\int f^-\,dp < \infty$, (43) reduces to the classical definition of $\int f\,dp$. We collect some of the properties of integration in this extended sense in the following lemma.

**Lemma 7.11** Let $X$ be a metrizable space and let $p \in P(X)$ be given. Let $f$, $g$ and $f_n$, $n = 1, 2, \ldots$, be Borel-measurable, extended real-valued functions on $X$.

(a) Using (42) to define $f + g$, we have

$$\int (f + g)\,dp \le \int f\,dp + \int g\,dp. \tag{44}$$

(b) If either
(b1) $\int f^+\,dp < \infty$ and $\int g^+\,dp < \infty$, or
(b2) $\int f^-\,dp < \infty$ and $\int g^-\,dp < \infty$, or
(b3) $\int g^+\,dp < \infty$ and $\int g^-\,dp < \infty$, then

$$\int (f + g)\,dp = \int f\,dp + \int g\,dp. \tag{45}$$

(c) If $0 < \alpha < \infty$, then $\int (\alpha f)\,dp = \alpha \int f\,dp$.
(d) If $f \le g$, then $\int f\,dp \le \int g\,dp$.
(e) If $f_n \uparrow f$ and $\int f_1\,dp > -\infty$, then $\int f_n\,dp \uparrow \int f\,dp$.
(f) If $f_n \downarrow f$ and $\int f_1\,dp < \infty$, then $\int f_n\,dp \downarrow \int f\,dp$.

*Proof* We prove (b) first and then return to (a). Under assumption (b1), we have $f(x) < \infty$ and $g(x) < \infty$ for $p$ almost every $x$, so the sum $f(x) + g(x)$ can be defined without resort to the convention (42) for $p$ almost every $x$. Furthermore, $\int f\,dp < \infty$ and $\int g\,dp < \infty$, so (45) follows from the additivity theorem for classical integration theory (Ash [A1, p. 45]). The proof of (45) under assumption (b2) is similar. Under assumption (b3), either $\int f^+\,dp = \infty$, in which case both sides of (45) are $\infty$, or else $\int f^+\,dp < \infty$, in which case assumption (b1) holds. Returning to (a), we note that if assumption (b1) holds, then (45) implies (44). If assumption (b1) fails to hold, then

$$\int f\,dp + \int g\,dp = \infty,$$

so (44) is still valid. Statements (c) and (d) are simple consequences of (42) and (43). Statement (e) follows from the extended monotone convergence theorem (Ash [A1, p. 47]) if $\int f_1^-\,dp < \infty$. If $\int f_1^-\,dp = \infty$, then $\int f_1\,dp > -\infty$ implies $\int f_1^+\,dp = \int f_1\,dp = \infty$, and the conclusion follows from (d). Statement (f) follows from the extended monotone convergence theorem. Q.E.D.

We saw in Corollary 7.27.2 that a probability measure on a product of Borel spaces can be decomposed into a stochastic kernel and a marginal. This process can be reversed, that is, given a probability measure and one or more Borel-measurable stochastic kernels on Borel spaces, a unique probability measure on the product space can be constructed.

**Proposition 7.28** Let $X_1, X_2, \ldots$ be a sequence of Borel spaces, $Y_n = X_1 X_2 \cdots X_n$ and $Y = X_1 X_2 \cdots$. Let $p \in P(X_1)$ be given, and, for $n = 1, 2, \ldots$, let $q_n(dx_{n+1}|y_n)$ be a Borel-measurable stochastic kernel on $X_{n+1}$ given $Y_n$.

Then for $n = 2, 3, \ldots$, there exist unique probability measures $r_n \in P(Y_n)$ such that

$$r_n(\underline{X}_1 \underline{X}_2 \cdots \underline{X}_n) = \int_{\underline{X}_1} \int_{\underline{X}_2} \cdots \int_{\underline{X}_{n-1}} q_{n-1}(\underline{X}_n | x_1, x_2, \ldots, x_{n-1}) \times q_{n-2}(dx_{n-1} | x_1, x_2, \ldots, x_{n-2}) \cdots \times q_1(dx_2 | x_1) p(dx_1) \qquad \forall \underline{X}_1 \in \mathscr{B}_{X_1}, \ldots, \underline{X}_n \in \mathscr{B}_{X_n}. \tag{46}$$

If $f: Y_n \to R^*$ is Borel-measurable and either $\int f^+ \, dr_n < \infty$ or $\int f^- \, dr_n < \infty$, then

$$\int_{Y_n} f \, dr_n = \int_{X_1} \int_{X_2} \cdots \int_{X_n} f(x_1, x_2, \ldots, x_n) q_{n-1}(dx_n | x_1, x_2, \ldots, x_{n-1}) \cdots \times q_1(dx_2 | x_1) p(dx_1). \tag{47}$$

Furthermore, there exists a unique probability measure $r$ on $Y = X_1 X_2 \cdots$ such that for each $n$ the marginal of $r$ on $Y_n$ is $r_n$.

*Proof* The spaces $Y_n$, $n = 2, 3, \ldots$, and $Y$ are Borel by Proposition 7.13. If there exists $r_n \in P(Y_n)$ satisfying (46), it must be unique. To see this, suppose $r_n' \in P(Y_n)$ also satisfies (46). The collection $\mathscr{D} = \{B \in \mathscr{B}_{Y_n} | r_n(B) = r_n'(B)\}$ is a Dynkin system containing the measurable rectangles, so $\mathscr{D} = \mathscr{B}_{Y_n}$ and $r_n = r_n'$. We establish the existence of $r_n$ by induction, considering first the case $n = 2$. For $B \in \mathscr{B}_{Y_2}$, use Corollary 7.26.1 to define

$$r_2(B) = \int_{X_1} q_1(B_{x_1} | x_1) p(dx_1). \tag{48}$$

It is easily verified that $r_2 \in P(Y_2)$ and $r_2$ satisfies (46). If $f$ is the indicator of $B \in \mathscr{B}_{Y_2}$, the $\int_{X_2} f(x_1, x_2) q_1(dx_2 | x_1)$ is Borel-measurable and, by (48),

$$\int_{Y_2} f \, dr_2 = \int_{X_1} \int_{X_2} f(x_1, x_2) q_1(dx_2 | x_1) p(dx_1). \tag{49}$$

Linearity of the integral implies that (49) holds for Borel-measurable simple functions as well. If $f: Y_2 \to [0, \infty]$ is Borel-measurable, then there exists an increasing sequence of simple functions such that $f_n \uparrow f$. By the monotone convergence theorem,

$$\lim_{n \to \infty} \int_{X_2} f_n(x_1, x_2) q_1(dx_2 | x_1) = \int_{X_2} f(x_1, x_2) q_1(dx_2 | x_1) \qquad \forall x_1 \in X_1,$$

so $\int_{X_2} f(x_1, x_2) q_1(dx_2 | x_1)$ is Borel-measurable and

$$\lim_{n \to \infty} \int_{Y_2} f_n \, dr_2 = \lim_{n \to \infty} \int_{X_1} \int_{X_2} f_n(x_1, x_2) q_1(dx_2 | x_1) p(dx_1) = \int_{X_1} \int_{X_2} f(x_1, x_2) q_1(dx_2 | x_1) p(dx_1).$$

But $\int_{Y_2} f_n \, dr_2 \uparrow \int_{Y_2} f \, dr_2$, so (49) holds for any Borel-measurable nonnegative $f$. For a Borel-measurable $f: Y_2 \to R^*$ satisfying $\int f^+ \, dr_n < \infty$ or $\int f^- \, dr_n < \infty$, we have

$$\int_{Y_2} f^+ \, dr_2 = \int_{X_1} \int_{X_2} f^+(x_1, x_2) q_1(dx_2|x_1) p(dx_1),$$

and

$$\int_{Y_2} f^- \, dr_2 = \int_{X_1} \int_{X_2} f^-(x_1, x_2) q_1(dx_2|x_1) p(dx_1).$$

Assume for specificity that $\int_{Y_2} f^- \, dr_2 < \infty$. Then the functions

$$\int_{X_2} f^+(x_1, x_2) q_1(dx_2|x_1)$$

and

$$-\int_{X_2} f^-(x_1, x_2) q_1(dx_2|x_1)$$

satisfy condition (b2) of Lemma 7.11, so

$$\begin{aligned}
\int_{Y_2} f \, dr_2 &= \int_{Y_2} f^+ \, dr_2 - \int_{Y_2} f^- \, dr_2 \\
&= \int_{X_1} \left[ \int_{X_2} f^+(x_1, x_2) q_1(dx_2|x_1) - \int_{X_2} f^-(x_1, x_2) q_1(dx_2|x_1) \right] p(dx_1) \\
&= \int_{X_1} \int_{X_2} f(x_1, x_2) q_1(dx_2|x_1) p(dx_1),
\end{aligned}$$

where the last step is a direct result of the definition of $\int_{X_2} f(x_1, x_2) q_1(dx_2|x_1)$.

Assume now that $r_k \in P(Y_k)$ exists for which (46) and (47) hold when $n = k$. For $B \in Y_{k+1}$, let

$$r_{k+1}(B) = \int_{Y_k} q_k(B_{y_k}|y_k) r_k(dy_k).$$

Then $r_{k+1} \in P(Y_{k+1})$. If $B = \underline{X}_1 \underline{X}_2 \cdots \underline{X}_k \underline{X}_{k+1}$, where $\underline{X}_j \in \mathscr{B}_{X_j}$, then

$$\begin{aligned}
r_{k+1}(B) &= \int \chi_{\underline{X}_1 \underline{X}_2 \cdots \underline{X}_k}(y_k) q_k(\underline{X}_{k+1}|y_k) r_k(dy_k) \\
&= \int_{\underline{X}_1} \int_{\underline{X}_2} \cdots \int_{\underline{X}_k} q_k(\underline{X}_{k+1}|x_1, x_2, \ldots, x_k) q_{k-1}(dx_k|x_{k-1}) \cdots \\
&\quad \times q_1(dx_2|x_1) p(dx_1) \qquad (50)
\end{aligned}$$

by (47) when $n = k$. This proves (46) for $n = k + 1$. Now use (50) to prove (47) when $n = k + 1$ and $f$ is an indicator function. As before, extend this to the case of $f: Y_{k+1} \to [0, \infty]$. If $f: Y_{k+1} \to R^*$ is Borel-measurable and either $\int f^+ \, dr_{k+1} < \infty$ or $\int f^- \, dr_{k+1} < \infty$, then the validity of (47) for nonnegative functions and the induction hypothesis imply

$$\int_{Y_{k+1}} f^+\, dr_{k+1} = \int_{X_1}\int_{X_2}\cdots\int_{X_{k+1}} f^+(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k)\cdots \times q_1(dx_2|x_1)p(dx_1)$$
$$= \int_{X_1\cdots X_k}\int_{X_{k+1}} f^+(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k)\, dr_k,$$

and likewise

$$\int_{Y_{k+1}} f^-\, dr_{k+1} = \int_{X_1\ldots X_k}\int_{X_{k+1}} f^-(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k)\, dr_k.$$

Assume for specificity that $\int_{Y_{k+1}} f^-\, dr_{k+1} < \infty$. Then the functions

$$\int_{X_{k+1}} f^+(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k)$$

and

$$-\int_{X_{k+1}} f^-(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k)$$

satisfy condition (b2) of Lemma 7.11, so as before

$$\int_{Y_{k+1}} f\, dr_{k+1} = \int_{X_1\cdots X_k}\int_{X_{k+1}} f(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k)\, dr_k. \tag{51}$$

Since

$$\left[\int_{X_{k+1}} f(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k)\right]^- \le \int_{X_{k+1}} f^-(x_1,\ldots,x_{k+1})q_k(dx_{k+1}|x_1,x_2,\ldots,x_k),$$

we can apply the induction hypothesis to the right-hand side of (51) to conclude that (47) holds in the generality stated in the proposition.

To establish the existence of a unique probability measure $r \in P(Y)$ whose marginal on $Y_n$ is $r_n$, $n = 2,3,\ldots$, we note that the measures $r_n$ are consistent, i.e., if $m \ge n$, then the marginal of $r_m$ on $Y_n$ is $r_n$. If each $X_k$ is complete, the Kolmogorov extension theorem (see, e.g., Ash [A1, p. 191]) guarantees the existence of a unique $r \in P(Y)$ whose marginal on each $Y_n$ is $r_n$. If $X_k$ is not complete, it can be homeomorphically embedded as a Borel subset in a complete separable metric space $\tilde{X}_k$. As in Proposition 7.13, each $Y_n$ is homeomorphic to a Borel subset of the complete separable metric space $\tilde{Y}_n = \tilde{X}_1\tilde{X}_2\cdots\tilde{X}_n$ and $Y$ is homeomorphic to a Borel subset of the complete separable metric space $\tilde{Y} = \tilde{X}_1\tilde{X}_2\cdots$. Each $r_n \in P(Y_n)$ can be identified with

$\tilde{r}_n \in P(\tilde{Y}_n)$ in the manner of Lemma 7.10, and, invoking the Kolmogorov extension theorem, we establish the existence of a unique $\tilde{r} \in P(\tilde{Y})$ whose marginal on each $\tilde{Y}_n$ is $\tilde{r}_n$. It is straightforward to show that $\tilde{r}$ assigns probability one to the image of $Y$ in $\tilde{Y}$, so $\tilde{r}$ corresponds to some $r \in P(Y)$ whose marginal on each $Y_n$ is $r_n$. The uniqueness of $\tilde{r}$ implies the uniqueness of $r$. Q.E.D.

In the course of proving Proposition 7.28, we have also proved the following.

**Proposition 7.29** Let $X$ and $Y$ be Borel spaces and $q(dy|x)$ a Borel-measurable stochastic kernel on $Y$ given $X$. If $f:XY \to R^*$ is Borel-measurable, then the function $\lambda:X \to R^*$ defined by

$$\lambda(x) = \int f(x, y)q(dy|x) \tag{52}$$

is Borel-measurable.

**Corollary 7.29.1** Let $X$ be a Borel space and let $f:X \to R^*$ be Borel-measurable. Then the function $\theta_f:P(X) \to R^*$ defined by

$$\theta_f(p) = \int f\, dp$$

is Borel-measurable.

*Proof* Define a Borel-measurable stochastic kernel on $X$ given $P(X)$ by $q(dx|p) = p(dx)$. Define $\tilde{f}:P(X)X \to R^*$ by $\tilde{f}(p, x) = f(x)$. Then

$$\theta_f(p) = \int f(x)p(dx) = \int \tilde{f}(p, x)q(dx|p)$$

is Borel-measurable by Proposition 7.29. Q.E.D.

If $f \in C(XY)$ and $q(dy|x)$ is continuous, then the mapping $\lambda$ of (52) is also continuous. We prove this with the aid of the following lemma.

**Lemma 7.12** Let $X$ and $Y$ be separable metrizable spaces. Then the mapping $\sigma:P(X)P(Y) \to P(XY)$ defined by

$$\sigma(p, q) = pq,$$

where $pq$ is the product of the measures $p$ and $q$, is continuous.

*Proof* We use Urysohn's theorem (Proposition 7.2) to homeomorphically embed $X$ and $Y$ into the Hilbert cube $\mathscr{H}$, and, for simplicity of notation, we treat $X$ and $Y$ as subsets of $\mathscr{H}$. Let $d$ be a metric on $\mathscr{H}\mathscr{H}$ consistent with its topology. If $g \in U_d(XY)$, then Lemma 7.3 implies that $g$ can be extended to a function $\hat{g} \in C(\mathscr{H}\mathscr{H})$. The set of finite linear combinations of the form $\sum_{j=1}^{k} \hat{f}_j(x)\hat{h}_j(y)$, where $\hat{f}_j$ and $\hat{h}_j$ range over $C(\mathscr{H})$ and $k$ ranges over the positive integers, is an algebra which separates points in $\mathscr{H}\mathscr{H}$, so given $\varepsilon > 0$, the Stone–Weierstrass theorem implies that such a linear combination can be found satisfying $\|\sum_{j=1}^{k} \hat{f}_j \hat{h}_j - \hat{g}\| < \varepsilon$. If $\{p_n\}$ is a sequence in $P(X)$ converging to $p \in P(X)$, $\{q_n\}$ a sequence in $P(Y)$ converging to $q \in P(Y)$, and $f_j$ and $h_j$ the restrictions of $\hat{f}_j$ and $\hat{h}_j$ to $X$ and $Y$, respectively, then

$$\begin{aligned}
&\limsup_{n\to\infty} \left| \int_{XY} g\, d(p_n q_n) - \int_{XY} g\, d(pq) \right| \\
&\quad \le \limsup_{n\to\infty} \left| \int_{XY} \left( g - \sum_{j=1}^{k} f_j h_j \right) d(p_n q_n) \right| \\
&\qquad + \sum_{j=1}^{k} \lim_{n\to\infty} \left| \int_X f_j\, dp_n \int_Y h_j\, dq_n - \int_X f_j\, dp \int_Y h_j\, dq \right| \\
&\qquad + \limsup_{n\to\infty} \left| \int_{XY} \left( \sum_{j=1}^{k} f_j h_j - g \right) d(pq) \right| \\
&\quad \le 2\varepsilon.
\end{aligned}$$

The continuity of $\sigma$ follows from the equivalence of (a) and (c) of Proposition 7.21. Q.E.D.

**Proposition 7.30** Let $X$ and $Y$ be separable metrizable spaces and let $q(dy|x)$ be a continuous stochastic kernel on $Y$ given $X$. If $f \in C(XY)$, then the function $\lambda: X \to R$ defined by

$$\lambda(x) = \int f(x, y) q(dy|x)$$

is continuous.

*Proof* The mapping $v: X \to P(XY)$ defined by $v(x) = p_x q(dy|x)$ is continuous by Corollary 7.21.1 and Lemma 7.12. We have $\lambda(x) = (\theta_f \circ v)(x)$, where $\theta_f: P(XY) \to R$ is defined by $\theta_f(r) = \int f\, dr$. By Proposition 7.21, $\theta_f$ is continuous. Hence, $\lambda$ is continuous. Q.E.D.

## 7.5 Semicontinuous Functions and Borel-Measurable Selection

In the dynamic programming algorithm given by (17) and (18) of Chapter 1, three operations are performed repetitively. First, there is the evaluation of a conditional expectation. Second, an extended real-valued function in two variables (state and control) is infimized over one of these variables (control). Finally, if an optimal or nearly optimal policy is to be constructed, a "selector" which maps each state to a control which achieves or nearly achieves the infimum in the second step must be chosen. In this section, we give results which will enable us to show that, under certain conditions, the extended real-valued functions involved are semicontinuous and the selectors can be chosen to be Borel-measurable. The results are applied to dynamic programming in Propositions 8.6–8.7 and Corollaries 9.17.2–9.17.3.

**Definition 7.13** Let $X$ be a metrizable space and $f$ an extended real-valued function on $X$. If $\{x \in X | f(x) \leq c\}$ is closed for every $c \in R$, $f$ is said to be *lower semicontinuous.* If $\{x \in X | f(x) \geq c\}$ is closed for every $c \in R$, $f$ is said to be *upper semicontinuous.*

Note that $f$ is lower semicontinuous if and only if $-f$ is upper semicontinuous. We will use this duality in the proofs of the following propositions to assert facts about upper semicontinuous functions given analogous facts about lower semicontinuous functions. Note also that if $f$ is lower semicontinuous, the sets $\{x \in X | f(x) = -\infty\}$ and $\{x \in X | f(x) \leq \infty\}$ are closed, since the former is equal to $\bigcap_{n=1}^{\infty}\{x \in X | f(x) \leq -n\}$ and the latter is $X$. There is a similar result for upper semicontinuous functions. The following lemma provides an alternative characterization of lower and upper semicontinuous functions.

**Lemma 7.13** Let $X$ be a metrizable space and $f: X \to R^*$.

(a) The function $f$ is lower semicontinuous if and only if for each sequence $\{x_n\} \subset X$ converging to $x \in X$

$$\liminf_{n\to\infty} f(x_n) \geq f(x). \tag{53}$$

(b) The function $f$ is upper semicontinuous if and only if for each sequence $\{x_n\} \subset X$ converging to $x \in X$

$$\limsup_{n\to\infty} f(x_n) \leq f(x). \tag{54}$$

*Proof* Suppose $f$ is lower semicontinuous and $x_n \to x$. We can extract a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to x$ as $k \to \infty$ and

$$\lim_{k\to\infty} f(x_{n_k}) = \liminf_{n\to\infty} f(x_n).$$

Given $\varepsilon > 0$, define

$$\theta(\varepsilon) = \begin{cases} \liminf\limits_{n\to\infty} f(x_n) + \varepsilon & \text{if } \liminf\limits_{n\to\infty} f(x_n) > -\infty, \\ -1/\varepsilon & \text{otherwise.} \end{cases}$$

There exists a positive integer $k(\varepsilon)$ such that $f(x_{n_k}) \le \theta(\varepsilon)$ for all $k \ge k(\varepsilon)$. The set $\{y \in X | f(y) \le \theta(\varepsilon)\}$ is closed, and hence it contains $x$. Inequality (53) follows. Conversely, if (53) holds and for some $c \in R$, $\{x_n\}$ is a sequence in $\{y \in X | f(y) \le c\}$ converging to $x$, then $f(x) \le c$, so $f$ is lower semicontinuous.

Part (b) of the proposition follows from part (a) by the duality mentioned earlier. Q.E.D.

If $f$ and $g$ are lower semicontinuous and bounded below on $X$ and if $x_n \to x$, then

$$\liminf_{n\to\infty}[f(x_n) + g(x_n)] \ge \liminf_{n\to\infty} f(x_n) + \liminf_{n\to\infty} g(x_n)$$
$$\ge f(x) + g(x),$$

so $f + g$ is lower semicontinuous. If $f$ is lower semicontinuous and $\alpha > 0$, then $\alpha f$ is lower semicontinuous as well. Upper semicontinuous functions have similar properties.

It is clear from (53) and (54) that $f: X \to R^*$ is continuous if and only if it is both lower and upper semicontinuous. We can often infer properties of semicontinuous functions from properties of continuous functions by means of the next lemma.

**Lemma 7.14** Let $X$ be a metrizable space and $f: X \to R^*$.

(a) The function $f$ is lower semicontinuous and bounded below if and only if there exists a sequence $\{f_n\} \subset C(X)$ such that $f_n \uparrow f$.

(b) The function $f$ is upper semicontinuous and bounded above if and only if there exists a sequence $\{f_n\} \subset C(X)$ such that $f_n \downarrow f$.

*Proof* We prove only part (a) of the proposition and appeal to duality for part (b). Assume $f$ is lower semicontinuous and bounded below by $b \in R$, and let $d$ be a metric on $X$ consistent with its topology. We may assume without loss of generality that for some $x_0 \in X$, $f(x_0) < \infty$, since the result is trivial otherwise. (Take $f_n(x) = n$ for every $x \in X$.) As in Lemma 7.7, define

$$g_n(x) = \inf_{y\in X}[f(y) + nd(x, y)].$$

Exactly as in the proof of Lemma 7.7, we show that $\{g_n\}$ is an increasing sequence of continuous functions bounded below by $b$ and above by $f$. The characterization (53) of lower semicontinuous functions can be used in place of continuity to prove $g_n \uparrow f$. In particular, (28) becomes

$$f(x) \le \liminf_{n\to\infty} f(y_n) \le \lim_{n\to\infty} g_n(x) + \varepsilon.$$

Now define $f_n = \min\{n, g_n\}$. Then each $f_n$ is continuous and bounded and $f_n \uparrow f$. This concludes the proof of the direct part of the proposition. For the converse part, suppose $\{f_n\} \subset C(X)$ and $f_n \uparrow f$. For $c \in R$,

$$\{x \in X | f(x) \le c\} = \bigcap_{n=1}^{\infty} \{x \in X | f_n(x) \le c\}$$

is closed. Q.E.D.

The following proposition shows that the semicontinuity of a function of two variables is preserved when one of the variables is integrated out via a continuous stochastic kernel.

**Proposition 7.31** Let $X$ and $Y$ be separable metrizable spaces, let $q(dy|x)$ be a continuous stochastic kernel on $Y$ given $X$, and let $f: XY \to R^*$ be Borel-measurable. Define

$$\lambda(x) = \int f(x, y) q(dy|x).$$

(a) If $f$ is lower semicontinuous and bounded below, then $\lambda$ is lower semicontinuous and bounded below.
(b) If $f$ is upper semicontinuous and bounded above, then $\lambda$ is upper semicontinuous and bounded above.

*Proof* We prove part (a) of the proposition and appeal to duality for part (b). If $f: XY \to R^*$ is lower semicontinuous and bounded below, then by Lemma 7.14 there exists a sequence $\{f_n\} \subset C(XY)$ such that $f_n \uparrow f$. Define $\lambda_n(x) = \int f_n(x, y) q(dy|x)$. By Proposition 7.30, we have that $\lambda_n$ is continuous, and by the monotone convergence theorem $\lambda_n \uparrow \lambda$. By Lemma 7.14, $\lambda$ is lower semicontinuous. Q.E.D.

An important operation in the execution of the dynamic programming algorithm is the infimization over one of the variables of a bivariate function. In the context of semicontinuity, we have the following result related to this operation.

**Proposition 7.32** Let $X$ and $Y$ be metrizable spaces and let $f: XY \to R^*$ be given. Define

$$f^*(x) = \inf_{y \in Y} f(x, y). \tag{55}$$

(a) If $f$ is lower semicontinuous and $Y$ is compact, then $f^*$ is lower semicontinuous and for every $x \in X$ the infimum in (55) is attained by some $y \in Y$.
(b) If $f$ is upper semicontinuous, then $f^*$ is upper semicontinuous.

*Proof* (a) Fix $x$ and let $\{y_n\} \subset Y$ be such that $f(x, y_n) \downarrow f^*(x)$. Then $\{y_n\}$ accumulates at some $y_0 \in Y$, and part (a) of Lemma 7.13 implies that $f(x, y_0) = f^*(x)$. To show that $f^*$ is lower semicontinuous, let $\{x_n\} \subset X$ be such that $x_n \to x_0$. Choose a sequence $\{y_n\} \subset Y$ such that

$$f(x_n, y_n) = f^*(x_n), \qquad n = 1, 2, \ldots.$$

There is a subsequence of $\{(x_n, y_n)\}$, call it $\{(x_{n_k}, y_{n_k})\}$, such that $\liminf_{n\to\infty} f(x_n, y_n) = \lim_{k\to\infty} f(x_{n_k}, y_{n_k})$. The sequence $\{y_{n_k}\}$ accumulates at some $y_0 \in Y$, and, by Lemma 7.13(a),

$$\liminf_{n\to\infty} f^*(x_n) = \liminf_{n\to\infty} f(x_n, y_n) = \lim_{k\to\infty} f(x_{n_k}, y_{n_k}) \geq f(x_0, y_0) \geq f^*(x_0),$$

so $f^*$ is lower semicontinuous.

(b) Let $d_1$ be a metric on $X$ and $d_2$ a metric on $Y$ consistent with their topologies. If $G \subset XY$ is open and $x_0 \in \mathrm{proj}_X(G)$, then there is some $y_0 \in Y$ for which $(x_0, y_0) \in G$, and there is some $\varepsilon > 0$ such that

$$N_\varepsilon(x_0, y_0) = \{(x, y) \in XY \mid d_1(x, x_0) < \varepsilon, d_2(y, y_0) < \varepsilon\} \subset G.$$

Then

$$x_0 \in \mathrm{proj}_X[N_\varepsilon(x_0, y_0)] = \{x \in X \mid d_1(x, x_0) < \varepsilon\} \subset \mathrm{proj}_X(G),$$

so $\mathrm{proj}_X(G)$ is open in $X$. For $c \in R$,

$$\{x \in X \mid f^*(x) < c\} = \mathrm{proj}_X(\{(x, y) \in XY \mid f(x, y) < c\}).$$

The upper semicontinuity of $f$ implies that $\{(x, y) \mid f(x, y) < c\}$ is open, so $\{x \in X \mid f^*(x) < c\}$ is open and $f^*$ is upper semicontinuous. Q.E.D.

Another important operation in the dynamic programming algorithm is the choice of a measurable "selector" which assigns to each $x \in X$ a $y \in Y$ which attains or nearly attains the infimum in (55). We first discuss Borel-measurable selection in case (a) of Proposition 7.32. For this we will need the Hausdorff metric and the corresponding topology on the set $2^Y$ of closed subsets of a compact metric space $Y$ (Appendix C). The space $2^Y$ under this topology is compact (Proposition C.2) and, therefore, complete and separable. Several preliminary lemmas are required.

**Lemma 7.15** Let $Y$ be a compact metrizable space and let $g: Y \to R^*$ be lower semicontinuous. Define $g^*: 2^Y \to R^*$ by

$$g^*(A) = \begin{cases} \min_{y \in A} g(y) & \text{if } A \neq \varnothing, \\ \infty & \text{if } A = \varnothing. \end{cases} \tag{56}$$

Then $g^*$ is lower semicontinuous.

*Proof* Since the empty set is an isolated point in $2^Y$, we need only prove that $g^*$ is lower semicontinuous on $2^Y - \{\varnothing\}$. We have already shown [Proposition 7.32(a)] that, given a nonempty set $A \in 2^Y$, there exists $y \in A$

such that $g^*(A) = g(y)$. Let $\{A_n\} \subset 2^Y$ be a sequence of nonempty sets with limit $A \in 2^Y$, and let $y_n \in A_n$ be such that $g^*(A_n) = g(y_n)$, $n = 1, 2, \ldots$. Choose a subsequence $\{y_{n_k}\}$ such that

$$\lim_{k\to\infty} g(y_{n_k}) = \liminf_{n\to\infty} g(y_n) = \liminf_{n\to\infty} g^*(A_n).$$

The subsequence $\{y_{n_k}\}$ accumulates at some $y_0 \in Y$, and, by Lemma 7.13(a),

$$g(y_0) \le \lim_{k\to\infty} g(y_{n_k}) = \liminf_{n\to\infty} g^*(A_n).$$

From (14) of Appendix C and from Proposition C.3, we have (in the notation of Appendix C)

$$y_0 \in \overline{\lim_{n\to\infty}}\, A_n = A,$$

so

$$g^*(A) \le g(y_0) \le \liminf_{n\to\infty} g^*(A_n).$$

The result follows from Lemma 7.13(a). Q.E.D.

**Lemma 7.16** Let $Y$ be a compact metrizable space and let $g: Y \to R^*$ be lower semicontinuous. Define $G: 2^Y R^* \to 2^Y$ by

$$G(A, c) = A \cap \{y \in Y \mid g(y) \le c\}. \tag{57}$$

Then $G$ is Borel-measurable.

*Proof* We show that $G$ is upper semicontinuous (K) (Definition C.2) and apply Proposition C.4. Let $\{(A_n, c_n)\} \subset 2^Y R^*$ be a sequence with limit $(A, c)$. If $\overline{\lim}_{n\to\infty} G(A_n, c_n) = \varnothing$, then

$$\overline{\lim_{n\to\infty}}\, G(A_n, c_n) \subset G(A, c). \tag{58}$$

Otherwise, choose $y \in \overline{\lim}_{n\to\infty} G(A_n, c_n)$. There is a sequence $n_1 < n_2 < \cdots$ of positive integers and a sequence $y_{n_k} \in G(A_{n_k}, c_{n_k})$, $k = 1, 2, \ldots$, such that $y_{n_k} \to y$. By definition, $y_{n_k} \in A_{n_k}$ for every $k$, so $y \in \overline{\lim}_{n\to\infty} A_n = A$. We also have $g(y_{n_k}) \le c_{n_k}$, $k = 1, 2, \ldots$, and using the lower semicontinuity of $g$, we obtain

$$g(y) \le \liminf_{k\to\infty} g(y_{n_k}) \le \lim_{k\to\infty} c_{n_k} = c.$$

Therefore $y \in G(A, c)$, (58) holds, and $G$ is upper semicontinuous (K).

Q.E.D.

**Lemma 7.17** Let $Y$ be a compact metrizable space and let $g: Y \to R^*$ be lower semicontinuous. Let $g^*: 2^Y \to R^*$ be defined by (56) and define $G^*: 2^Y \to 2^Y$ by

$$G^*(A) = A \cap \{y \in Y | g(y) \le g^*(A)\}. \tag{59}$$

Then $G^*$ is Borel-measurable.

*Proof* Let $G$ be the Borel-measurable function given by (57). Lemma 7.15 implies $g^*$ is Borel-measurable. A comparison of (57) and (59) shows that

$$G^*(A) = G[A, g^*(A)].$$

It follows that $G^*$ is also Borel-measurable. Q.E.D.

**Lemma 7.18** Let $Y$ be a compact metrizable space. There is a Borel-measurable function $\sigma: 2^Y - \{\varnothing\} \to Y$ such that $\sigma(A) \in A$ for every $A \in 2^Y - \{\varnothing\}$.

*Proof* Let $\{g_n | n = 1, 2, \ldots\}$ be a subset of $C(Y)$ which separates points in $Y$ (for example, the one constructed in the proof of Proposition 7.7). As in Lemma 7.15, define $g_n^*: 2^Y \to R^*$ by

$$g_n^*(A) = \begin{cases} \min_{y \in A} g_n(y) & \text{if } A \neq \varnothing, \\ \infty & \text{if } A = \varnothing, \end{cases}$$

and, as in Lemma 7.17, define $G_n^*: 2^Y \to 2^Y$ by

$$G_n^*(A) = A \cap \{y \in Y | g_n(y) \le g_n^*(A)\} = \{y \in A | g_n(y) = g_n^*(A)\}.$$

Let $H_n: 2^Y \to 2^Y$ be defined recursively by

$$H_0(A) = A,$$
$$H_n(A) = G_n^*[H_{n-1}(A)], \qquad n = 1, 2, \ldots.$$

Then for $A \neq \varnothing$, each $H_n(A)$ is nonempty and compact, and

$$A = H_0(A) \supset H_1(A) \supset H_2(A) \supset \cdots.$$

Therefore, $\bigcap_{n=0}^{\infty} H_n(A) \neq \varnothing$. If $y, y' \in \bigcap_{n=0}^{\infty} H_n(A)$, then for $n = 1, 2, \ldots$, we have

$$g_n(y) = g_n^*[H_{n-1}(A)] = g_n(y').$$

Since $\{g_n | n = 1, 2, \ldots\}$ separates points in $Y$, we have $y = y'$, and $\bigcap_{n=0}^{\infty} H_n(A)$ must consist of a single point, which we denote by $\sigma(A)$.

We show that for $A \neq \varnothing$

$$\lim_{n\to\infty} H_n(A) = \bigcap_{n=0}^{\infty} H_n(A) = \{\sigma(A)\}. \tag{60}$$

Since the sequence $\{H_n(A)\}$ is nonincreasing, we have from (14) and (15) of Appendix C that

$$\bigcap_{n=0}^{\infty} H_n(A) \subset \underline{\lim}_{n\to\infty} H_n(A) \subset \overline{\lim}_{n\to\infty} H_n(A). \tag{61}$$

If $y \in \overline{\lim}_{n\to\infty} H_n(A)$, then there exist positive integers $n_1 < n_2 < \cdots$ and a sequence $y_{n_k} \in H_{n_k}(A)$, $k = 1, 2, \ldots$, such that $y_{n_k} \to y$. For fixed $k$, $y_{n_j} \in H_{n_k}$ for all $j \geq k$, and since $H_{n_k}(A)$ is closed, we have $y \in H_{n_k}(A)$. Therefore, $y \in \bigcap_{n=0}^{\infty} H_n(A)$ and

$$\overline{\lim}_{n\to\infty} H_n(A) \subset \bigcap_{n=0}^{\infty} H_n(A). \tag{62}$$

From relations (61) and (62), we obtain (60).

Since $G_n^*$ and $H_n$ are Borel-measurable for every $n$, the mapping $v: 2^Y - \{\varnothing\} \to 2^Y$ defined by $v(A) = \{\sigma(A)\}$ is Borel-measurable. It is easily seen that the mapping $\tau: Y \to 2^Y$ defined by $\tau(y) = \{y\}$ is a homeomorphism. Since $Y$ is compact, $\tau(Y)$ is compact, thus closed in $2^Y$, and $\tau^{-1}: \tau(Y) \to Y$ is Borel-measurable. Since $\sigma = \tau^{-1} \circ v$, it follows that $\sigma$ is Borel-measurable.

Q.E.D.

**Lemma 7.19** Let $X$ be a metrizable space, $Y$ a compact metrizable space, and let $f: XY \to R^*$ be lower semicontinuous. Define $F: XR^* \to 2^Y$ by

$$F(x, c) = \{y \in Y \mid f(x, y) \leq c\}. \tag{63}$$

Then $F$ is Borel-measurable.

*Proof* The proof is very similar to that of Lemma 7.16. We show that $F$ is upper semicontinuous $(K)$ and apply Proposition C.4. Let $(x_n, c_n) \to (x, c)$ in $XR^*$ and let $y$ be an element of $\overline{\lim}_{n\to\infty} F(x_n, c_n)$, provided this set is nonempty. There exist positive integers $n_1 < n_2 < \cdots$ and $y_{n_k} \in F(x_{n_k}, c_{n_k})$ such that $y_{n_k} \to y$. Since $f(x_{n_k}, y_{n_k}) \leq c_{n_k}$ and $f$ is lower semicontinuous, we conclude that $f(x, y) \leq c$, so that $\overline{\lim}_{n\to\infty} F(x_n, c_n) \subset F(x, c)$. The result follows.

Q.E.D.

**Lemma 7.20** Let $X$ be a metrizable space, $Y$ a compact metrizable space, and let $f: XY \to R^*$ be lower semicontinuous. Let $f^*: X \to R^*$ be given by $f^*(x) = \min_{y \in Y} f(x, y)$, and define $F^*: X \to 2^Y$ by

$$F^*(x) = \{y \in Y \mid f(x, y) \leq f^*(x)\}. \tag{64}$$

Then $F^*$ is Borel-measurable.

*Proof* Let $F$ be the Borel-measurable function defined by (63). Proposition 7.32(a) implies that $f^*$ is Borel-measurable. From (63) and (64) we have

$$F^*(x) = F[x, f^*(x)].$$

It follows that $F^*$ is also Borel-measurable. Q.E.D.

We are now ready to prove the selection theorem for lower semicontinuous functions.

**Proposition 7.33** Let $X$ be a metrizable space, $Y$ a compact metrizable space, $D$ a closed subset of $XY$, and let $f: D \to R^*$ be lower semicontinuous. Let $f^*: \mathrm{proj}_X(D) \to R^*$ be given by

$$f^*(x) = \min_{y \in D_x} f(x, y). \tag{65}$$

Then $\mathrm{proj}_X(D)$ is closed in $X$, $f^*$ is lower semicontinuous, and there exists a Borel-measurable function $\varphi: \mathrm{proj}_X(D) \to Y$ such that $\mathrm{Gr}(\varphi) \subset D$ and

$$f[x, \varphi(x)] = f^*(x) \qquad \forall x \in \mathrm{proj}_X(D). \tag{66}$$

*Proof* We first prove the result for the case where $D = XY$. As in Lemma 7.18, let $\sigma: 2^Y - \{\varnothing\} \to Y$ be a Borel-measurable function satisfying $\sigma(A) \in A$ for every $A \in 2^Y - \{\varnothing\}$. As in Lemma 7.20, let $F^*: X \to 2^Y$ be the Borel-measurable function defined by

$$F^*(x) = \{y \in Y | f(x, y) = f^*(x)\}.$$

Proposition 7.32(a) implies that $f^*$ is lower semicontinuous and $F^*(x) \neq \varnothing$ for every $x \in X$. The composition $\varphi = \sigma \circ F^*$ satisfies (66).

Suppose now that $D$ is not necessarily $XY$. To see that $\mathrm{proj}_X(D)$ is closed, note that the function $g = -\chi_D$ is lower semicontinuous and

$$\mathrm{proj}_X(D) = \{x \in X | g^*(x) \leq -1\},$$

where $g^*(x) = \min_{y \in Y} g(x, y)$. By the special case of the proposition already proved, $g^*$ is lower semicontinuous, $\mathrm{proj}_X(D)$ is closed, and there is a Borel-measurable function $\varphi_1: X \to Y$ such that $g[x, \varphi_1(x)] = g^*(x)$ for every $x \in X$ or, equivalently, $(x, \varphi_1(x)) \in D$ for every $x \in \mathrm{proj}_X(D)$.

Define now the lower semicontinuous function $\hat{f}: XY \to R^*$ by

$$\hat{f}(x, y) = \begin{cases} f(x, y) & \text{if } (x, y) \in D, \\ \infty & \text{otherwise.} \end{cases}$$

For all $c \in R$,

$$\{x \in \mathrm{proj}_X(D) | f^*(x) \leq c\} = \left\{x \in X \Big| \min_{y \in Y} \hat{f}(x, y) \leq c\right\}.$$

Since $\min_{y \in Y} \hat{f}(x, y)$ is lower semicontinuous, it follows that $f^*$ is also lower semicontinuous. Let $\varphi_2 : X \to Y$ be a Borel-measurable function satisfying

$$\hat{f}[x, \varphi_2(x)] = \min_{y \in Y} \hat{f}(x, y) \qquad \forall x \in X.$$

Clearly $(x, \varphi_2(x)) \in D$ for all $x$ in the Borel set

$$\left\{ x \in X \middle| \min_{y \in Y} \hat{f}(x, y) < \infty \right\}.$$

Define $\varphi : \mathrm{proj}_X(D) \to Y$ by

$$\varphi(x) = \begin{cases} \varphi_1(x) & \text{if } \min_{y \in Y} \hat{f}(x, y) = \infty, \\ \varphi_2(x) & \text{if } \min_{y \in Y} \hat{f}(x, y) < \infty. \end{cases}$$

The function $\varphi$ is Borel-measurable and satisfies (66). Q.E.D.

We turn our attention to selection in the case of an upper semicontinuous function. The analysis is considerably simpler, but in contrast to the "exact selector" of (66) we will obtain only an approximate selector for this case.

**Lemma 7.21** Let $X$ be a metrizable space, $Y$ a separable metrizable space, and $G$ an open subset of $XY$. Then $\mathrm{proj}_X(G)$ is open and there exists a Borel-measurable function $\varphi : \mathrm{proj}_X(G) \to Y$ such that $\mathrm{Gr}(\varphi) \subset G$.

*Proof* Let $\{ y_n | n = 1, 2, \ldots \}$ be a countable dense subset of $Y$. For fixed $y \in Y$, the mapping $x \to (x, y)$ is continuous, so $\{x \in X | (x, y) \in G\}$ is open. Let $G_n = \{x \in X | (x, y_n) \in G\}$, and note that $\mathrm{proj}_X(G) = \bigcup_{n=1}^{\infty} G_n$ is open. Define $\varphi : \mathrm{proj}_X(G) \to Y$ by

$$\varphi(x) = \begin{cases} y_1 & \text{if } x \in G_1, \\ y_n & \text{if } x \in G_n - \bigcup_{k=1}^{n-1} G_k, \quad n = 2, 3, \ldots. \end{cases}$$

Then $\varphi$ is Borel-measurable and $\mathrm{Gr}(\varphi) \subset G$. Q.E.D.

**Proposition 7.34** Let $X$ be a metrizable space, $Y$ a separable metrizable space, $D$ an open subset of $XY$, and let $f : D \to R^*$ be upper semicontinuous. Let $f^* : \mathrm{proj}_X(D) \to R^*$ be given by

$$f^*(x) = \inf_{y \in D_x} f(x, y). \tag{67}$$

Then $\mathrm{proj}_X(D)$ is open in $X$, $f^*$ is upper semicontinuous, and for every $\varepsilon > 0$, there exists a Borel-measurable function $\varphi_\varepsilon : \mathrm{proj}_X(D) \to Y$ such that

$\mathrm{Gr}(\varphi_\varepsilon) \subset D$ and for all $x \in \mathrm{proj}_X(D)$

$$f[x, \varphi_\varepsilon(x)] \le \begin{cases} f^*(x) + \varepsilon & \text{if} \quad f^*(x) > -\infty, \\ -1/\varepsilon & \text{if} \quad f^*(x) = -\infty. \end{cases} \tag{68}$$

*Proof* The set $\mathrm{proj}_X(D)$ is open in $X$ by Lemma 7.21. To show that $f^*$ is upper semicontinuous, define an upper semicontinuous function $\hat{f}: XY \to R^*$ by

$$\hat{f}(x, y) = \begin{cases} f(x, y) & \text{if} \quad (x, y) \in D, \\ \infty & \text{otherwise.} \end{cases}$$

For $c \in R$, we have

$$\{x \in \mathrm{proj}_X(D) | f^*(x) < c\} = \left\{x \in X \Big| \inf_{y \in Y} \hat{f}(x, y) < c\right\},$$

and this set is open by Proposition 7.32(b).

Let $\varepsilon > 0$ be given. For $k = 0, \pm 1, \pm 2, \ldots$, define (see Fig. 7.1)

$$\begin{aligned} A(k) &= \{(x, y) \in D | f(x, y) < k\varepsilon\}, \\ B(k) &= \{x \in \mathrm{proj}_X(D) | (k-1)\varepsilon \le f^*(x) < k\varepsilon\}, \\ B(-\infty) &= \{x \in \mathrm{proj}_X(D) | f^*(x) = -\infty\}, \\ B(\infty) &= \{x \in \mathrm{proj}_X(D) | f^*(x) = \infty\}. \end{aligned}$$

The sets $A(k)$, $k = 0, \pm 1, \pm 2, \ldots$, are open, while the sets $B(k)$, $k = 0, \pm 1, \pm 2, \ldots, B(-\infty)$, and $B(\infty)$ are Borel-measurable. By Lemma 7.21,

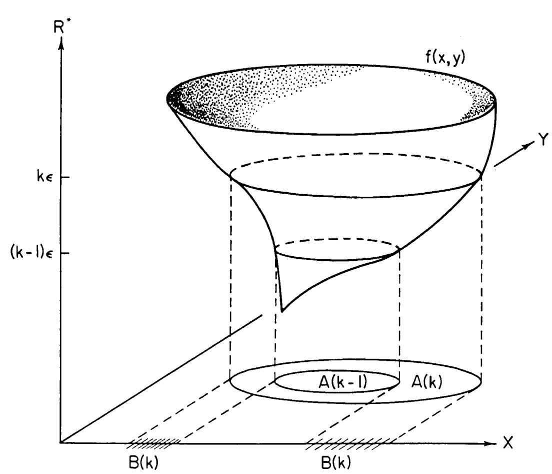


FIGURE 7.1

there exists for each $k = 0, \pm 1, \pm 2, \ldots$ a Borel-measurable $\varphi_k : \mathrm{proj}_X(A_k) \to Y$ such that $\mathrm{Gr}(\varphi_k) \subset A_k$, and there exists a Borel-measurable $\bar{\varphi} : \mathrm{proj}_X(D) \to Y$ such that $\mathrm{Gr}(\bar{\varphi}) \subset D$. Let $k^*$ be an integer such that $k^* \leq -1/\varepsilon^2$. Define $\varphi_\varepsilon : \mathrm{proj}_X(D) \to Y$ by

$$\varphi_\varepsilon(x) = \begin{cases} \varphi_k(x) & \text{if } x \in B(k), \quad k = 0, \pm 1, \pm 2, \ldots, \\ \bar{\varphi}(x) & \text{if } x \in B(\infty), \\ \varphi_{k^*}(x) & \text{if } x \in B(-\infty). \end{cases}$$

Since $B(k) \subset \mathrm{proj}_X[A(k)]$ and $B(-\infty) \subset \mathrm{proj}_X[A(k)]$ for all $k$, this definition is possible. It is clear that $\varphi_\varepsilon$ is Borel-measurable and $\mathrm{Gr}(\varphi_\varepsilon) \subset D$. If $x \in B(k)$, then, since $(x, \varphi_k(x)) \in A(k)$, we have

$$f[x, \varphi_\varepsilon(x)] = f[x, \varphi_k(x)] < k\varepsilon \leq f^*(x) + \varepsilon.$$

If $x \in B(\infty)$, then $f(x, y) = \infty$ for all $y \in D_x$ and $f[x, \varphi_\varepsilon(x)] = \infty = f^*(x)$. If $x \in B(-\infty)$, we have

$$f[x, \varphi_\varepsilon(x)] = f[x, \varphi_{k^*}(x)] < k^*\varepsilon \leq -1/\varepsilon.$$

Hence $\varphi_\varepsilon$ has the required properties. Q.E.D.

## 7.6 Analytic Sets

The dynamic programming algorithm is centered around infimization of functions, and this is intimately connected with projections of sets. More specifically, if $f : XY \to R^*$ is given and $f^* : X \to R^*$ is defined by

$$f^*(x) = \inf_{y \in Y} f(x, y),$$

then for each $c \in R$

$$\{x \in X \mid f^*(x) < c\} = \mathrm{proj}_X(\{(x, y) \in XY \mid f(x, y) < c\}).$$

If $f$ is a Borel-measurable function, then $\{(x, y) \mid f(x, y) < c\}$ is a Borel-measurable set. Unfortunately, the projection of a Borel-measurable set need not be Borel-measurable. In Borel spaces, however, the projection of a Borel set is an analytic set. This section is devoted to development of properties of analytic sets.

### *7.6.1 Equivalent Definitions of Analytic Sets*

There are a number of ways to define the class of analytic sets in a Borel space $X$. One possibility is to define them as the projections on $X$ of the Borel subsets of $XY$, where $Y$ is some uncountable Borel space. Another possibility is to define them as the images of the Baire null space $\mathcal{N}$ under continuous functions from $\mathcal{N}$ into $X$. Still another possibility is to define them as all sets of the form

$$\bigcup_{(\sigma_1,\sigma_2,\ldots)\in\mathcal{N}} \bigcap_{n=1}^{\infty} S(\sigma_1,\sigma_2,\ldots,\sigma_n),$$

where $\mathcal{N}$ is the set of all sequences of positive integers (the Baire null space) and the sets $S(\sigma_1,\sigma_2,\ldots,\sigma_n)$ are closed in $X$. All these definitions are equivalent, as we show in Proposition 7.41. We will take the third definition as our starting point, since this is the most convenient analytically. We first formalize the set operation just given in terms of the notion of a Suslin scheme in a paved space.

**Definition 7.14** Let $X$ be a set. A *paving* $\mathscr{P}$ of $X$ is a nonempty collection of subsets of $X$. The pair $(X,\mathscr{P})$ is called a *paved space.*

If $(X,\mathscr{P})$ is a paved space, we denote by $\sigma(\mathscr{P})$ the $\sigma$-algebra generated by $\mathscr{P}$, we denote by $\mathscr{P}_\delta$ the collection of all intersections of countably many members of $\mathscr{P}$, and we denote by $\mathscr{P}_\sigma$ the collection of all unions of countably many members of $\mathscr{P}$. Recall that $N$ is the set of positive integers, $\mathcal{N}$ is the set of all infinite sequences of positive integers, and $\Sigma$ is the set of all finite sequences of positive integers.

**Definition 7.15** Let $(X,\mathscr{P})$ be a paved space. A *Suslin scheme* for $\mathscr{P}$ is a mapping from $\Sigma$ into $\mathscr{P}$. The *nucleus of a Suslin scheme* $S:\Sigma\to\mathscr{P}$ is

$$N(S) = \bigcup_{(\sigma_1,\sigma_2,\ldots)\in\mathcal{N}} \bigcap_{n=1}^{\infty} S(\sigma_1,\sigma_2,\ldots,\sigma_n). \tag{69}$$

The set of all nuclei of Suslin schemes for a paving $\mathscr{P}$ will be denoted by $\mathscr{S}(\mathscr{P})$.

In order to simplify notation, we write, for $s=(\sigma_1,\sigma_2,\ldots,\sigma_n)\in\Sigma$ and $z=(\zeta_1,\zeta_2,\ldots)\in\mathcal{N}$,

$$s<z \qquad \text{if} \quad \sigma_1=\zeta_1, \quad \sigma_2=\zeta_2,\ldots,\sigma_n=\zeta_n.$$

With this notation, (69) can also be written as

$$N(S) = \bigcup_{z\in\mathcal{N}} \bigcap_{s<z} S(s).$$

We will use both expressions interchangeably.

Note that the union in (69) is uncountable, so if $\mathscr{P}$ is a $\sigma$-algebra and $S$ is a Suslin scheme for $\mathscr{P}$, $N(S)$ may be outside $\mathscr{P}$. Several properties of $\mathscr{S}(\mathscr{P})$ are given below.

**Proposition 7.35** Let $X$ be a space with pavings $\mathscr{P}$ and $\mathscr{Q}$ such that $\mathscr{P} \subset \mathscr{Q}$. Then

(a) $\mathscr{S}(\mathscr{P}) \subset \mathscr{S}(\mathscr{Q})$,
(b) $\mathscr{S}(\mathscr{P})_\delta = \mathscr{S}(\mathscr{P})$,
(c) $\mathscr{S}(\mathscr{P})_\sigma = \mathscr{S}(\mathscr{P})$,
(d) $\mathscr{P} \subset \mathscr{S}(\mathscr{P})$,
(e) $\mathscr{S}(\mathscr{P}) = \mathscr{S}[\mathscr{S}(\mathscr{P})]$.

*Proof* (a) Obvious.

(b) It is clear that $\mathscr{S}(\mathscr{P})_\delta \supset \mathscr{S}(\mathscr{P})$. Now choose $\bigcap_{k=1}^{\infty} N(S_k) \in \mathscr{S}(\mathscr{P})_\delta$, where $S_k$ is a Suslin scheme for $\mathscr{P}$, $k = 1, 2, \ldots$. It suffices to construct a Suslin scheme $S$ for $\mathscr{P}$ such that

$$N(S) = \bigcap_{k=1}^{\infty} N(S_k). \tag{70}$$

For $k = 1, 2, \ldots$, let $\Pi_k = \{(2j-1)2^{k-1} \mid j = 1, 2, \ldots\}$. Then $\Pi_1, \Pi_2, \ldots$ is a partition of $N$ into infinitely many infinite sets. For each positive integer $k$, let $\varphi_k : \mathscr{N} \to \mathscr{N}$ be defined by

$$\varphi_k(\zeta_1, \zeta_2, \ldots) = (\zeta_{2^{k-1}}, \zeta_{3\cdot 2^{k-1}}, \zeta_{5\cdot 2^{k-1}}, \ldots),$$

i.e., $\varphi_k$ picks out the components of $(\zeta_1, \zeta_2, \ldots)$ with indices in $\Pi_k$. We want to construct a Suslin scheme $S$ for which

$$\bigcap_{s<z} S(s) = \bigcap_{k=1}^{\infty} \bigcap_{s<\varphi_k(z)} S_k(s) \qquad \forall z \in \mathscr{N}. \tag{71}$$

We may rewrite (71) as

$$\bigcap_{n=1}^{\infty} S(\zeta_1, \zeta_2, \ldots, \zeta_n)$$
$$= \bigcap_{k=1}^{\infty} \bigcap_{j=1}^{\infty} S_k(\zeta_{2^{k-1}}, \zeta_{3\cdot 2^{k-1}}, \ldots, \zeta_{(2j-1)2^{k-1}}) \qquad \forall (\zeta_1, \zeta_2, \ldots) \in \mathscr{N}. \tag{72}$$

Given $(\zeta_1, \zeta_2, \ldots, \zeta_n) \in \Sigma$, we have $n = (2j-1)2^{k-1}$ for exactly one pair of positive integers $j$ and $k$. Define

$$S(\zeta_1, \zeta_2, \ldots, \zeta_n) = S_k(\zeta_{2^{k-1}}, \zeta_{3\cdot 2^{k-1}}, \ldots, \zeta_{(2j-1)2^{k-1}}). \tag{73}$$

This defines a Suslin scheme $S$ for which (72), and hence (71), is easily verified.

We now use (71) to prove (70). Choose

$$x \in N(S) = \bigcup_{z \in \mathscr{N}} \bigcap_{s<z} S(s).$$

For some $z_0 \in \mathcal{N}$, we have

$$x \in \bigcap_{s < z_0} S(s) = \bigcap_{k=1}^{\infty} \bigcap_{s < \varphi_k(z_0)} S_k(s).$$

Thus, for every $k$,

$$x \in \bigcap_{s < \varphi_k(z_0)} S_k(s) \subset \bigcup_{z \in \mathcal{N}} \bigcap_{s < z} S_k(s) = N(S_k).$$

It follows that $x \in \bigcap_{k=1}^{\infty} N(S_k)$ and

$$N(S) \subset \bigcap_{k=1}^{\infty} N(S_k). \tag{74}$$

If we are given $x \in \bigcap_{k=1}^{\infty} N(S_k)$, then $x \in \bigcup_{z \in \mathcal{N}} \bigcap_{s<z} S_k(s)$ for every $k$, so for every $k$, there exists $z_k \in \mathcal{N}$ such that $x \in \bigcap_{s < z_k} S_k(s)$. Let $z_0 \in \mathcal{N}$ be such that $\varphi_k(z_0) = z_k$, $k = 1, 2, \ldots$. Then $x \in \bigcap_{k=1}^{\infty} \bigcap_{s < \varphi_k(z_0)} S_k(s)$. An application of (71) shows that

$$x \in \bigcap_{s < z_0} S(s) \subset \bigcup_{z \in \mathcal{N}} \bigcap_{s < z} S(s) = N(S)$$

and

$$N(S) \supset \bigcap_{k=1}^{\infty} N(S_k). \tag{75}$$

Relation (70) follows from (74) and (75).

(c) It is clear that $\mathcal{S}(\mathcal{P})_\sigma \supset \mathcal{S}(\mathcal{P})$. Choose $\bigcup_{k=1}^{\infty} N(S_k) \in \mathcal{S}(\mathcal{P})_\sigma$, where $S_k$ is a Suslin scheme for $\mathcal{P}$, $k = 1, 2, \ldots$. It suffices to construct a Suslin scheme $S$ for $\mathcal{P}$ for which

$$N(S) = \bigcup_{k=1}^{\infty} N(S_k). \tag{76}$$

Given $(\zeta_1, \zeta_2, \ldots, \zeta_n) \in \Sigma$, we have $\zeta_1 = (2j-1)2^{k-1}$ for exactly one pair of positive integers $j$ and $k$. Define

$$S(\zeta_1, \zeta_2, \ldots, \zeta_n) = S((2j-1)2^{k-1}, \zeta_2, \ldots, \zeta_n) = S_k(j, \zeta_2, \ldots, \zeta_n).$$

This defines a Suslin scheme $S$ for which

$$\bigcap_{n=1}^{\infty} S((2j-1)2^{k-1}, \zeta_2, \ldots, \zeta_n) = \bigcap_{n=1}^{\infty} S_k(j, \zeta_2, \ldots, \zeta_n)$$
$$\forall k \in N, \forall (j, \zeta_2, \ldots) \in \mathcal{N}. \tag{77}$$

Returning to (76), we choose $x \in N(S) = \bigcup_{z \in \mathcal{N}} \bigcap_{s<z} S(s)$. For some $(\zeta_1, \zeta_2, \ldots) \in \mathcal{N}$, we have $x \in \bigcap_{n=1}^{\infty} S(\zeta_1, \zeta_2, \ldots, \zeta_n)$, and choosing $j$ and $k$ so

that $\zeta_1 = (2j-1)2^{k-1}$, we have, from (77),

$$x \in \bigcap_{n=1}^{\infty} S_k(j, \zeta_2, \ldots, \zeta_n) \subset N(S_k) \subset \bigcup_{k=1}^{\infty} N(S_k),$$

so

$$N(S) \subset \bigcup_{k=1}^{\infty} N(S_k). \tag{78}$$

If, on the other hand, we choose

$$x \in \bigcup_{k=1}^{\infty} N(S_k) = \bigcup_{k=1}^{\infty} \bigcup_{z \in \mathcal{N}} \bigcap_{s < z} S_k(s),$$

then for some $k \in N$ and $(j, \zeta_2, \ldots) \in \mathcal{N}$, we have $x \in \bigcap_{n=1}^{\infty} S_k(j, \zeta_2, \ldots, \zeta_n)$. Equation (77) implies

$$x \in \bigcap_{n=1}^{\infty} S((2j-1)2^{k-1}, \zeta_2, \ldots, \zeta_n) \subset N(S),$$

so

$$N(S) \supset \bigcup_{k=1}^{\infty} N(S_k). \tag{79}$$

Relation (76) follows from (78) and (79).

(d) For $P \in \mathcal{P}$, define $S(s) = P$ for every $s \in \Sigma$. Then $N(S) = P$.

(e) The proof of this takes us somewhat far afield, so is given as Proposition B.2 of Appendix B. Q.E.D.

It is not in general true that $\mathcal{S}(\mathcal{P})$ is closed under complementation, so $\mathcal{S}(\mathcal{P})$ is generally not a $\sigma$-algebra. In order for $\mathcal{S}(\mathcal{P})$ to contain $\sigma(\mathcal{P})$, we need one additional assumption.

**Corollary 7.35.1** Let $(X, \mathcal{P})$ be a paved space and assume that the complement of each set in $\mathcal{P}$ is in $\mathcal{S}(\mathcal{P})$. Then $\sigma(\mathcal{P}) \subset \mathcal{S}(\mathcal{P})$.

*Proof* The smallest algebra containing $\mathcal{P}$ consists of the finite intersections of finite unions of sets in $\mathcal{P}$ and complements of sets in $\mathcal{P}$. By Proposition 7.35, these sets are contained in $\mathcal{S}[\mathcal{S}[\mathcal{S}(\mathcal{P})]] = \mathcal{S}(\mathcal{P})$. Since $\mathcal{S}(\mathcal{P})$ is a monotone class, it contains the $\sigma$-algebra generated by $\mathcal{P}$ as well (Ash [A1, p. 19]). Q.E.D.

**Definition 7.16** Let $X$ be a Borel space. Denote by $\mathcal{F}_X$ the collection of closed subsets of $X$. The *analytic subsets* of $X$ are the members of $\mathcal{S}(\mathcal{F}_X)$.

**Corollary 7.35.2** Let $X$ be a Borel space. The countable intersections and unions of analytic subsets of $X$ are analytic.

*Proof* This follows from Proposition 7.35(b) and (c). Q.E.D.

**Proposition 7.36** Let $X$ be a Borel space. Then every Borel subset of $X$ is analytic. Indeed, the class of analytic sets $\mathscr{S}(\mathscr{F}_X)$ is equal to $\mathscr{S}(\mathscr{B}_X)$.

*Proof* Every open subset of $X$ is an $F_\sigma$ (Lemma 7.2), so every open set is analytic. Corollary 7.35.1 implies $\mathscr{B}_X \subset \mathscr{S}(\mathscr{F}_X)$. Proposition 7.35(a) and (e) implies

$$\mathscr{S}(\mathscr{F}_X) \subset \mathscr{S}(\mathscr{B}_X) \subset \mathscr{S}[\mathscr{S}(\mathscr{F}_X)] = \mathscr{S}(\mathscr{F}_X). \qquad \text{Q.E.D.}$$

If the Borel space $X$ is countable, then every subset of $X$ is both analytic and Borel-measurable. If $X$ is uncountable, however, the class of analytic subsets of $X$ is strictly larger than $\mathscr{B}_X$. This is shown in Appendix B, where we prove the existence of an analytic set whose complement is not analytic.

Note that an immediate consequence of Proposition 7.36 is that if $Y$ is a Borel subset of the Borel space $X$, then the analytic subsets of $Y$ are the analytic subsets of $X$ contained in $Y$. A generalization of this fact is the following.

**Corollary 7.36.1** Let $X$ and $Y$ be Borel spaces and $\varphi : X \to Y$ a Borel isomorphism. Then $A \subset X$ is analytic if and only if $\varphi(A) \subset Y$ is analytic.

*Proof* If $\varphi : X \to Y$ is a Borel isomorphism and $A \subset X$ is analytic, then $A = N(S)$, where $S$ is a Suslin scheme for $\mathscr{F}_X$. It is easily seen that $\varphi(A) = N(\varphi \circ S)$, where $\varphi \circ S$ is the Suslin scheme for $\mathscr{B}_Y$ defined by $(\varphi \circ S)(s) = \varphi[S(s)]$, so $\varphi(A)$ is analytic by Proposition 7.36. If $\varphi(A) \subset Y$ is analytic, $A \subset X$ is analytic by a similar argument. Q.E.D.

We proceed to the development of several equivalent characterizations of analytic sets. The general definition of a Suslin scheme is unrestrictive with respect to the form of the mapping $S : \Sigma \to \mathscr{P}$. In the event that $X$ is a separable metric space and $\mathscr{P} = \mathscr{F}_X$, one can assume without loss of generality that $S$ has more structure.

**Definition 7.17** Let $(X, \mathscr{P})$ be a paved space and $S$ a Suslin scheme for $\mathscr{P}$. The Suslin scheme $S$ is *regular* if for each $n \in N$ and $(\sigma_1, \sigma_2, \ldots, \sigma_{n+1}) \in \Sigma$, we have

$$S(\sigma_1, \sigma_2, \ldots, \sigma_n) \supset S(\sigma_1, \sigma_2, \ldots, \sigma_n, \sigma_{n+1}).$$

**Lemma 7.22** Let $(X, d)$ be a separable metric space and $S$ a Suslin scheme for $\mathscr{F}_X$. Then there exists a regular Suslin scheme $R$ for $\mathscr{F}_X$ such that $N(R) = N(S)$ and, for every $z = (\zeta_1, \zeta_2, \ldots) \in \mathscr{N}$,

$$\lim_{n \to \infty} \operatorname{diam} R(\zeta_1, \zeta_2, \ldots, \zeta_n) = 0 \quad \text{if} \quad R(\zeta_1, \zeta_2, \ldots, \zeta_n) \neq \varnothing \quad \forall n. \qquad (80)$$

*Proof* By the Lindelöf property, for each positive integer $k$, $X$ can be covered by a countable collection of open spheres of the form $B_{kj} = \{x\in X \mid d(x,x_{kj}) < 1/k\}$, $j = 1,2,\ldots$. For $(\bar\zeta_1,\zeta_1,\bar\zeta_2,\zeta_2,\ldots)\in\mathcal{N}$, define

$$\begin{aligned} R(\bar\zeta_1) &= \bar B_{1\bar\zeta_1},\\ R(\bar\zeta_1,\zeta_1) &= R(\bar\zeta_1)\cap S(\zeta_1),\\ R(\bar\zeta_1,\zeta_1,\bar\zeta_2) &= R(\bar\zeta_1,\zeta_1)\cap \bar B_{2\bar\zeta_2},\\ R(\bar\zeta_1,\zeta_1,\bar\zeta_2,\zeta_2) &= R(\bar\zeta_1,\zeta_1,\bar\zeta_2)\cap S(\zeta_1,\zeta_2), \end{aligned}$$

etc. Thus

$$\bigcap_{s<(\bar\zeta_1,\zeta_1,\bar\zeta_2,\zeta_2,\ldots)} R(s) = \left[\bigcap_{k=1}^{\infty} \bar B_{k\bar\zeta_k}\right] \cap \left[\bigcap_{s<z} S(s)\right], \tag{81}$$

where $z = (\zeta_1,\zeta_2,\ldots)$. It is clear that $R$ is a regular Suslin scheme for $\mathscr{F}_X$ and (80) holds. If $x\in N(R)$, then there exists $(\bar\zeta_1,\zeta_1,\bar\zeta_2,\zeta_2,\ldots)\in\mathcal{N}$ such that $x\in\bigcap_{s<(\bar\zeta_1,\zeta_1,\bar\zeta_2,\zeta_2,\ldots)} R(s)$, so by (81) $x\in\bigcap_{s<(\zeta_1,\zeta_2,\ldots)} S(s)\subset N(S)$, and therefore $N(R)\subset N(S)$. If $x\in N(S)$, then there exists $(\zeta_1,\zeta_2,\ldots)\in\mathcal{N}$ such that $x\in\bigcap_{s<(\zeta_1,\zeta_2,\ldots)} S(s)$. Since for each positive integer $k$, the collection $\{B_{kj} \mid j = 1,2,\ldots\}$ covers $X$, there exists for each $k$ a positive integer $\bar\zeta_k$ for which $x\in B_{k\bar\zeta_k}$. Then $x\in\bigcap_{k=1}^{\infty} B_{k\bar\zeta_k}$ and, by (81), $x\in\bigcap_{s<(\bar\zeta_1,\zeta_1,\bar\zeta_2,\zeta_2,\ldots)} R(s)\subset N(R)$, so $N(R)\supset N(S)$. It follows that $N(R) = N(S)$. Q.E.D.

Note that if a regular Suslin scheme $R$ satisfies (80), then for all $z$ in

$$\mathcal{N}_1 = \left\{ z\in\mathcal{N} \,\middle|\, \bigcap_{s<z} R(s) \neq \varnothing \right\},$$

the set $\bigcap_{s<z} R(s)$ consists of a single point, say $f(z)$. Thus we have

$$N(R) = f(\mathcal{N}_1),$$

and this relation provides the basis for an alternative way of characterizing analytic sets. We have the following lemma.

**Lemma 7.23** Let $(X,d)$ be a complete separable metric space. If $A\subset X$ is a nonempty analytic set, then there exist a closed subset $\mathcal{N}_1$ of $\mathcal{N}$ and a continuous function $f:\mathcal{N}_1\to X$ such that $A = f(\mathcal{N}_1)$. Conversely, if $\mathcal{N}_1\subset\mathcal{N}$ is closed and $f:\mathcal{N}_1\to X$ is continuous, then $f(\mathcal{N}_1)$ is analytic.

*Proof* Let $A = N(R)$ be nonempty, where $R$ is a regular Suslin scheme for $\mathscr{F}_X$ satisfying (80). Define

$$\mathcal{N}_1 = \left\{ z\in\mathcal{N} \,\middle|\, \bigcap_{s<z} R(s) \neq \varnothing \right\}.$$

Let $z = (\zeta_1, \zeta_2, \ldots)$ be in $\mathcal{N}$. If for each $n$ we have $R(\zeta_1, \zeta_2, \ldots, \zeta_n) \neq \varnothing$, then it is possible to chose $x_n \in R(\zeta_1, \zeta_2, \ldots, \zeta_n)$. The sequence $\{x_n\}$ is Cauchy by (80), and since $(X, d)$ is complete, $\{x_n\}$ has a limit $x \in X$. But for each $n$ the regularity of $R$ implies $\{x_m | m \geq n\} \subset R(\zeta_1, \zeta_2, \ldots, \zeta_n)$, so $x \in R(\zeta_1, \zeta_2, \ldots, \zeta_n)$. Therefore $x \in \bigcap_{s<z} R(s)$. Now suppose $z \in \mathcal{N} - \mathcal{N}_1$. The preceding argument shows that for some $s_n < z$, we have $R(s_n) = \varnothing$. The open neighborhood $\{w \in \mathcal{N} | s_n < w\}$ contains $z$ and is contained in $\mathcal{N} - \mathcal{N}_1$, so $\mathcal{N} - \mathcal{N}_1$ is open and $\mathcal{N}_1$ is closed.

For $z \in \mathcal{N}_1$, define $f(z)$ to be the unique point in $\bigcap_{s<z} R(s)$. If $\{z_k\}$ is a sequence in $\mathcal{N}_1$ converging to $z_0 = (\zeta_1, \zeta_2, \ldots) \in \mathcal{N}_1$, then given $\varepsilon > 0$, (80) implies that there exists $s_n < z_0$ for which $\operatorname{diam} R(s_n) < \varepsilon$. For $k$ sufficiently large, $z_k \in \{z \in \mathcal{N} | s_n < z\}$, so $f(z_k) \in R(s_n)$. Therefore $d(f(z_k), f(z_0)) \leq \operatorname{diam} R(s_n) < \varepsilon$, which shows that $f$ is continuous.

For the converse, suppose $\mathcal{N}_1 \subset \mathcal{N}$ is closed and $f: \mathcal{N}_1 \to X$ is continuous. Define a regular Suslin scheme $R$ for $\mathscr{F}_X$ by

$$R(s) = \overline{f(\{z \in \mathcal{N}_1 | s < z\})},$$

where $R(s) = \varnothing$ if $\{z \in \mathcal{N}_1 | s < z\} = \varnothing$. If $z \in \mathcal{N}_1$, then $f(z) \in \bigcap_{s<z} R(s) \subset N(R)$, so $f(\mathcal{N}_1) \subset N(R)$. If $x \in N(R)$, then for some $z_0 = (\zeta_1, \zeta_2, \ldots) \in \mathcal{N}$ we have $x \in \bigcap_{s<z_0} R(s)$. Then for each $n$,

$$x \in \overline{f(\{z \in \mathcal{N}_1 | (\zeta_1, \zeta_2, \ldots, \zeta_n) < z\})},$$

so given $\varepsilon > 0$, there exists a $z_n \in \mathcal{N}_1$ with $(\zeta_1, \zeta_2, \ldots, \zeta_n) < z_n$ and $d(x, f(z_n)) < \varepsilon$. But as $n \to \infty$, $z_n$ must converge to $z_0$. The closedness of $\mathcal{N}_1$ implies $z_0 \in \mathcal{N}_1$, and the continuity of $f$ implies $d(x, f(z_0)) \leq \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we have $f(z_0) = x$, $x \in f(\mathcal{N}_1)$, and $N(R) \subset f(\mathcal{N}_1)$. Q.E.D.

We have thus characterized analytic sets as the continuous images of closed subsets of $\mathcal{N}$. We will obtain an even sharper characterization, for which we need the following lemma.

**Lemma 7.24** If $\mathcal{N}_1$ is a nonempty closed subset of $\mathcal{N}$, then there exists a continuous function $g: \mathcal{N} \to \mathcal{N}$ such that $\mathcal{N}_1 = g(\mathcal{N})$.

*Proof* Use the Lindelöf property to cover $\mathcal{N}_1$ with a countable collection of nonempty closed sets $\{S(\zeta_1) | \zeta_1 \in N\}$ which satisfy

$$\mathcal{N}_1 \supset S(\zeta_1), \qquad \operatorname{diam} S(\zeta_1) \leq 1, \quad \zeta_1 = 1, 2, \ldots,$$

where $d$ is a metric on $\mathcal{N}$ consistent with its topology and $\operatorname{diam} S(\zeta_1)$ is given by (9). Cover each $S(\zeta_1)$ with a countable collection of nonempty closed sets $\{S(\zeta_1, \zeta_2) | \zeta_2 \in N\}$ which satisfy

$$S(\zeta_1) \supset S(\zeta_1, \zeta_2), \qquad \operatorname{diam} S(\zeta_1, \zeta_2) \leq \tfrac{1}{2}, \quad \zeta_2 = 1, 2, \ldots.$$

Continue in this manner so that, for any $(\zeta_1, \zeta_2, \ldots, \zeta_{n-1})$,

$$S(\zeta_1, \zeta_2, \ldots, \zeta_n) \neq \varnothing, \qquad \zeta_n = 1, 2, \ldots, \tag{82}$$

$$S(\zeta_1, \zeta_2, \ldots, \zeta_{n-1}) = \bigcup_{\zeta_n = 1}^{\infty} S(\zeta_1, \zeta_2, \ldots, \zeta_n), \tag{83}$$

$$S(\zeta_1, \zeta_2, \ldots, \zeta_{n-1}) \supset S(\zeta_1, \zeta_2, \ldots, \zeta_{n-1}, \zeta_n), \qquad \zeta_n = 1, 2, \ldots, \tag{84}$$

$$\operatorname{diam} S(\zeta_1, \zeta_2, \ldots, \zeta_n) \leq 1/n, \qquad \zeta_n = 1, 2, \ldots. \tag{85}$$

The completeness of $\mathcal{N}$ and (82)–(85) imply that for each $z \in \mathcal{N}$, $\bigcap_{s<z} S(s)$ consists of a single point. Define $g(z)$ to be this point. Then

$$N(S) = g(\mathcal{N}) = \mathcal{N}_1.$$

The continuity of $g$ follows by an argument similar to the one used in the proof of Lemma 7.23. Q.E.D.

**Proposition 7.37** Let $X$ be a Borel space. A nonempty set $A \subset X$ is analytic if and only if $A = f(\mathcal{N})$ for some continuous function $f: \mathcal{N} \to X$.

*Proof* If $X$ is complete, the proposition follows from Lemmas 7.23 and 7.24. If $X$ is not complete, it is still homeomorphic to a Borel subset of a complete separable space, and the result follows from Corollary 7.36.1. Q.E.D.

Proposition 7.37 gives a very useful characterization of nonempty analytic sets in terms of continuous functions and the Baire null space $\mathcal{N}$. The Baire null space has a simple description and its topology allows considerable flexibility. We have already shown, for example, that it is homeomorphic to $\mathcal{N}_0$, the space of irrationals in $[0, 1]$. Another important homeomorphism is the following.

**Lemma 7.25** The space $\mathcal{N}$ is homeomorphic to any finite or countably infinite product of copies of itself.

*Proof* We prove the lemma for the case of a countably infinite product. Let $\Pi_1, \Pi_2, \ldots$ be a partition of $N$, the set of positive integers, into infinitely many infinite sets. Define $\varphi: \mathcal{N} \to \mathcal{N}\mathcal{N}\mathcal{N} \cdots$ by

$$\varphi(z) = (z_1, z_2, \ldots), \tag{86}$$

where $z_k$ consists of the components of $z$ with indices in $\Pi_k$. Then $\varphi$ is one-to-one and onto and, because convergence in a product space is componentwise, $\varphi$ is a homeomorphism. Q.E.D.

Combination of Lemma 7.25 with Proposition 7.37 gives the following.

**Proposition 7.38** Let $X_1, X_2, \ldots$ be a sequence of Borel spaces and $A_k$ an analytic subset of $X_k$, $k = 1, 2, \ldots$. Then the sets $A_1 A_2 \cdots$ and $A_1 A_2 \cdots A_n$, $n = 1, 2, \ldots$, are analytic subsets of $X_1 X_2 \cdots$ and $X_1 X_2 \cdots X_n$, respectively.

*Proof* Let $f_k : \mathcal{N} \to X_k$ be continuous such that $A_k = f_k(\mathcal{N})$, $k = 1, 2, \ldots$. Let $\varphi$ be given by (86) and $F : \mathcal{N}\mathcal{N} \cdots \to X_1 X_2 \cdots$ be given by

$$F(z_1, z_2, \ldots) = (f_1(z_1), f_2(z_2), \ldots).$$

Then $F \circ \varphi$ is continuous and maps $\mathcal{N}$ onto $A_1 A_2 \cdots$. The finite products are handled similarly. Q.E.D.

Another consequence of Proposition 7.37 is that the continuous image of an analytic set, in particular, the projection of an analytic set, is analytic. As discussed at the beginning of this section, this property motivated our inquiry into analytic sets. We formalize this and a related fact to obtain another characterization of analytic sets.

**Proposition 7.39** Let $X$ and $Y$ be Borel spaces and $A$ an analytic subset of $XY$. Then $\mathrm{proj}_X(A)$ is analytic. Conversely, given any analytic set $C \subset X$ and any uncountable Borel space $Y$, there is a Borel set $B \subset XY$ such that $C = \mathrm{proj}_X(B)$. If $Y = \mathcal{N}$, $B$ can be chosen to be closed.

*Proof* If $A = f(\mathcal{N}) \subset XY$ is analytic, where $f$ is continuous, then $\mathrm{proj}_X(A) = (\mathrm{proj}_X \circ f)(\mathcal{N})$ is analytic by Proposition 7.37. If $C = f(\mathcal{N}) \subset X$ is nonempty and analytic, then

$$C = \mathrm{proj}_X[\widetilde{\mathrm{Gr}}(f)],$$

where $\widetilde{\mathrm{Gr}}(f) = \{(f(z), z) \in X\mathcal{N} \mid z \in \mathcal{N}\}$ is closed because $f$ is continuous. If $Y$ is any uncountable Borel space, then there exists a Borel isomorphism $\varphi$ from $\mathcal{N}$ onto $Y$ (Corollary 7.16.1). The mapping $\Phi$ defined by

$$\Phi(x, z) = (x, \varphi(z))$$

is a Borel isomorphism from $X\mathcal{N}$ onto $XY$, and

$$C = \mathrm{proj}_X(\Phi[\widetilde{\mathrm{Gr}}(f)]). \qquad \text{Q.E.D.}$$

So far we have treated only the continuous images of analytic sets. With the aid of Proposition 7.39, we can consider their images under Borel-measurable functions as well.

**Proposition 7.40** Let $X$ and $Y$ be Borel spaces and $f : X \to Y$ a Borel-measurable function. If $A \subset X$ is analytic, then $f(A)$ is analytic. If $B \subset Y$ is analytic, then $f^{-1}(B)$ is analytic.

*Proof* Suppose $A \subset X$ is analytic. By Proposition 7.39, there exists a Borel set $B \subset X\mathcal{N}$ such that $A = \mathrm{proj}_X(B)$. Define $\psi : B \to Y$ by $\psi(x, z) = f(x)$.

Then $\psi$ is Borel-measurable, and Corollary 7.14.1 implies that $\mathrm{Gr}(\psi) \in \mathscr{B}_{X\mathscr{N}Y}$. Finally, $f(A) = \mathrm{proj}_Y[\mathrm{Gr}(\psi)]$ is analytic by Proposition 7.39.

If $B \subset Y$ is analytic, then $B = N(S)$, where $S$ is some Suslin scheme for $\mathscr{F}_Y$. Then $f^{-1}(B) = N(f^{-1} \circ S)$, where $f^{-1} \circ S$ is the Suslin scheme for $\mathscr{B}_X$ defined by

$$(f^{-1} \circ S)(s) = f^{-1}[S(s)] \qquad \forall s \in \Sigma.$$

The analyticity of $f^{-1}(B)$ follows from Proposition 7.36. Q.E.D.

We summarize the equivalent definitions of analytic sets in Borel spaces.

**Proposition 7.41** Let $X$ be a Borel space. The following definitions of the collection of analytic subsets of $X$ are equivalent:

(a) $\mathscr{S}(\mathscr{F}_X)$;
(b) $\mathscr{S}(\mathscr{B}_X)$;
(c) the empty set and the images of $\mathscr{N}$ under continuous functions from $\mathscr{N}$ into $X$;
(d) the projections into $X$ of the closed subsets of $X\mathscr{N}$;
(e) the projections into $X$ of the Borel subsets of $XY$, where $Y$ is an uncountable Borel space;
(f) the images of Borel subsets of $Y$ under Borel-measurable functions from $Y$ into $X$, where $Y$ is an uncountable Borel space.

*Proof* The only new characterization here is $(f)$. If $Y$ is an uncountable Borel space and $f: Y \to X$ is Borel-measurable, then for every $B \in \mathscr{B}_Y$, $f(B)$ is analytic in $X$ by Proposition 7.40. To show that every nonempty analytic set $A \subset X$ can be obtained this way, let $\varphi$ be a Borel isomorphism from $Y$ onto $X\mathscr{N}$ and let $F \subset X\mathscr{N}$ be closed and satisfy $\mathrm{proj}_X(F) = A$. Define $B = \varphi^{-1}(F) \in \mathscr{B}_Y$. Then $(\mathrm{proj}_X \circ \varphi)(B) = A$. If $A = \varnothing$, then $f(\varnothing) = A$ for any Borel-measurable $f: Y \to X$. Q.E.D.

### *7.6.2 Measurability Properties of Analytic Sets*

At the beginning of this section we indicated that extended real-valued functions on a Borel space $X$ whose lower level sets are analytic arise naturally via partial infimization. Because the collection of analytic subsets of an uncountable Borel space is strictly larger than the Borel $\sigma$-algebra (Appendix B), such functions need not be Borel-measurable. Nonetheless, they can be integrated with respect to any probability measure on $(X, \mathscr{B}_X)$. To show this, we must discuss the measurability properties of analytic sets.

If $X$ is a Borel space and $p \in P(X)$, we define *p-outer measure*, denoted by $p^*$, on the set of all subsets of $X$ by

$$p^*(E) = \inf\{p(B) | E \subset B, B \in \mathscr{B}_X\}. \tag{87}$$

Outer measure on an increasing sequence of sets has a convergence property, namely, $p^*(E_n) \uparrow p^*(\bigcup_{n=1}^{\infty} E_n)$ if $E_1 \subset E_2 \subset \cdots$. This is easy to verify from (87) and also follows from Eq. (5) and Proposition A.1 of Appendix A (see also Ash [A1, Lemma 1.3.3(d)]). The collection of sets $\mathscr{B}_X(p)$ defined by

$$\mathscr{B}_X(p) = \{E \subset X \mid p^*(E) + p^*(E^c) = 1\}$$

is a $\sigma$-algebra (Ash [A1, Theorem 1.3.5]), called the *completion of* $\mathscr{B}_X$ *with respect to* $p$. It can be described as the class of sets of the form $B \cup N$ as $B$ ranges over $\mathscr{B}_X$ and $N$ ranges over all subsets of sets of $p$-measure zero in $\mathscr{B}_X$ (Ash [A1, p. 18]), and we have

$$p^*(B \cup N) = p(B).$$

Furthermore, $p^*$ restricted to $\mathscr{B}_X(p)$ is a probability measure, and is the only extension of $p$ to $\mathscr{B}_X(p)$ that is a probability measure. In what follows, we denote this measure also by $p$ and write $p(E)$ in place of $p^*(E)$ for all $E \in \mathscr{B}_X(p)$.

**Definition 7.18** Let $X$ be a Borel space. The *universal* $\sigma$*-algebra* $\mathscr{U}_X$ is defined by $\mathscr{U}_X = \bigcap_{p \in P(X)} \mathscr{B}_X(p)$. If $E \in \mathscr{U}_X$, we say $E$ is *universally measurable.*

The usefulness of analytic sets in measure theory is in large degree derived from the following proposition.

**Proposition 7.42** (Lusin's theorem) Let $X$ be a Borel space and $S$ a Suslin scheme for $\mathscr{U}_X$. Then $N(S)$ is universally measurable. In other words, $\mathscr{S}(\mathscr{U}_X) = \mathscr{U}_X$.

*Proof* Denote $A = N(S)$, where $S$ is a Suslin scheme for $\mathscr{U}_X$. For $(\sigma_1, \ldots, \sigma_k) \in \Sigma$, define

$$N(\sigma_1, \ldots, \sigma_k) = \{(\zeta_1, \zeta_2, \ldots) \in \mathscr{N} \mid \zeta_1 = \sigma_1, \ldots, \zeta_k = \sigma_k\} \tag{88}$$

and

$$\begin{aligned} M(\sigma_1, \ldots, \sigma_k) &= \{(\zeta_1, \zeta_2, \ldots) \in \mathscr{N} \mid \zeta_1 \le \sigma_1, \ldots, \zeta_k \le \sigma_k\} \\ &= \bigcup_{\tau_1 \le \sigma_1, \ldots, \tau_k \le \sigma_k} N(\tau_1, \ldots, \tau_k). \end{aligned} \tag{89}$$

Define also

$$R(\sigma_1, \ldots, \sigma_k) = \bigcup_{z \in M(\sigma_1, \ldots, \sigma_k)} \bigcap_{s < z} S(s). \tag{90}$$

Then

$$R(\sigma_1, \ldots, \sigma_k) \subset K(\sigma_1, \ldots, \sigma_k), \tag{91}$$

where

$$K(\sigma_1, \ldots, \sigma_k) = \bigcup_{\tau_1 \le \sigma_1, \ldots, \tau_k \le \sigma_k} \bigcap_{j=1}^{k} S(\tau_1, \ldots, \tau_j). \tag{92}$$

As $\sigma_1 \uparrow \infty$, $M(\sigma_1) \uparrow \mathcal{N}$, so $R(\sigma_1) \uparrow A$. Likewise, as $\sigma_k \uparrow \infty$, $M(\sigma_1, \ldots, \sigma_{k-1}, \sigma_k) \uparrow M(\sigma_1, \ldots, \sigma_{k-1})$, so $R(\sigma_1, \ldots, \sigma_{k-1}, \sigma_k) \uparrow R(\sigma_1, \ldots, \sigma_{k-1})$. Given $p \in P(X)$ and $\varepsilon > 0$, choose $\bar{\zeta}_1, \bar{\zeta}_2, \ldots$ such that

$$p^*(A) \le p^*[R(\bar{\zeta}_1)] + (\varepsilon/2),$$
$$p^*[R(\bar{\zeta}_1, \ldots, \bar{\zeta}_{k-1})] \le p^*[R(\bar{\zeta}_1, \ldots, \bar{\zeta}_{k-1}, \bar{\zeta}_k)] + (\varepsilon/2^k), \qquad k = 2, 3, \ldots.$$

Then

$$p^*(A) \le p^*[R(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)] + \varepsilon, \qquad k = 1, 2, \ldots. \tag{93}$$

The set $K(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)$ is universally measurable, so (91) and (93) imply

$$\begin{aligned} 1 &= p[K(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)] + p[X - K(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)] \\ &\ge p^*[R(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)] + p[X - K(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)] \\ &\ge p^*(A) - \varepsilon + p[X - K(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)]. \end{aligned} \tag{94}$$

We show that

$$\bigcap_{k=1}^{\infty} K(\zeta_1, \ldots, \zeta_k) \subset A \qquad \forall (\zeta_1, \zeta_2, \ldots) \in \mathcal{N}. \tag{95}$$

If

$$x \in \bigcap_{k=1}^{\infty} K(\zeta_1, \ldots, \zeta_k) = \bigcap_{k=1}^{\infty} \bigcup_{\tau_1 \le \zeta_1, \ldots, \tau_k \le \zeta_k} \bigcap_{j=1}^{k} S(\tau_1, \ldots, \tau_j), \tag{96}$$

an argument by contradiction will be used to show that for some $\bar{\tau}_1 \le \zeta_1$, we have

$$x \in S(\bar{\tau}_1) \cap \left[ \bigcap_{k=2}^{\infty} \bigcup_{\tau_2 \le \zeta_2, \ldots, \tau_k \le \zeta_k} \bigcap_{j=2}^{k} S(\bar{\tau}_1, \tau_2, \ldots, \tau_j) \right]. \tag{97}$$

If no such $\bar{\tau}_1$ existed, then for every $\tau_1 \le \zeta_1$, there would exist a positive integer $k(\tau_1)$ such that

$$x \notin S(\tau_1) \cap \left[ \bigcap_{k=2}^{k(\tau_1)} \bigcup_{\tau_2 \le \zeta_2, \ldots, \tau_k \le \zeta_k} \bigcap_{j=2}^{k} S(\tau_1, \tau_2, \ldots, \tau_j) \right].$$

If $\bar{k} = \max_{\tau_1 \le \zeta_1} k(\tau_1)$, then

$$\begin{aligned} x &\notin \bigcup_{\tau_1 \le \zeta_1} \left\{ S(\tau_1) \cap \left[ \bigcap_{k=2}^{\bar{k}} \bigcup_{\tau_2 \le \zeta_2, \ldots, \tau_k \le \zeta_k} \bigcap_{j=2}^{k} S(\tau_1, \tau_2, \ldots, \tau_j) \right] \right\} \\ &\supset \bigcup_{\tau_1 \le \zeta_1, \ldots, \tau_{\bar{k}} \le \zeta_{\bar{k}}} \bigcap_{j=1}^{\bar{k}} S(\tau_1, \tau_2, \ldots, \tau_j) = K(\zeta_1, \ldots, \zeta_{\bar{k}}) \end{aligned}$$

and a contradiction is reached. Replace (96) by (97) and apply the same argument to establish the existence of $\overline{\tau}_2 \leq \zeta_2$ such that

$$x \in S(\overline{\tau}_1) \cap S(\overline{\tau}_1, \overline{\tau}_2) \cap \left[ \bigcap_{k=3}^{\infty} \bigcup_{\tau_3 \leq \zeta_3, \ldots, \tau_k \leq \zeta_k} \bigcap_{j=3}^{k} S(\overline{\tau}_1, \overline{\tau}_2, \tau_3, \ldots, \tau_j) \right].$$

Continuing this process, construct a sequence $\overline{\tau}_1 \leq \zeta_1, \overline{\tau}_2 \leq \zeta_2, \ldots$ such that

$$x \in \bigcap_{k=1}^{\infty} S(\overline{\tau}_1, \ldots, \overline{\tau}_k) \subset N(S) = A.$$

This proves (95), i.e., as $k \to \infty$, $K(\zeta_1, \ldots, \zeta_k)$ decreases to a set contained in $A$, and $X - K(\zeta_1, \ldots, \zeta_k)$ increases to a set containing $X - A$. Letting $k \to \infty$ in (94), we obtain

$$1 \geq p^*(A) - \varepsilon + p^*(X - A).$$

Since $\varepsilon > 0$ is arbitrary, this implies that

$$1 \geq p^*(A) + p^*(X - A).$$

It is true for any $E \subset X$ that $p^*(E) + p^*(X - E) \geq 1$, so

$$p^*(A) + p^*(X - A) = 1,$$

and $A$ is measurable with respect to $p$. Q.E.D.

**Corollary 7.42.1** Let $X$ be a Borel space. Every analytic subset of $X$ is universally measurable.

*Proof* The closed subsets of $X$ are universally measurable, so $\mathscr{S}(\mathscr{F}_X) \subset \mathscr{U}_X$ by Proposition 7.42. Q.E.D.

As remarked earlier, the class of analytic subsets of an uncountable Borel space is not a $\sigma$-algebra, so there are universally measurable sets which are not analytic. In fact, we show in Appendix B that in any uncountable Borel space, the universal $\sigma$-algebra is strictly larger than the $\sigma$-algebra generated by the analytic subsets.

### *7.6.3 An Analytic Set of Probability Measures*

In Proposition 7.25 we saw that when $X$ is a Borel space, the function $\theta_A: P(X) \to [0, 1]$ defined by $\theta_A(p) = p(A)$ is Borel-measurable for every Borel-measurable $A \subset X$. We now investigate the properties of this function when $A$ is analytic. The main result is that the set $\{p \in P(X) | p(A) \geq c\}$ is analytic for each real $c$.

**Proposition 7.43** Let $X$ be a Borel space and $A$ an analytic subset of $X$. For each $c \in R$, the set $\{p \in P(X) | p(A) \geq c\}$ is analytic.

*Proof* Let $S$ be a Suslin scheme for $\mathscr{F}_X$, the class of closed subsets of $X$, such that $A = N(S)$. For $s \in \Sigma$, let $N(s)$, $M(s)$, $R(s)$, and $K(s)$ be defined by (88)–(90) and (92). Then (91) and (95) hold and each $K(s)$ is closed. We show that for $c \in R$

$$\{p \in P(X) | p(A) \geq c\} = \bigcap_{n=1}^{\infty} \bigcup_{z \in \mathscr{N}} \bigcap_{s < z} \{p \in P(X) | p[K(s)] \geq c - (1/n)\}. \quad (98)$$

If $\overline{p}(A) \geq c$, then for any $n \geq 1$, there exists $(\overline{\zeta}_1, \overline{\zeta}_2, \ldots) \in \mathscr{N}$ such that (93) is satisfied with $p = \overline{p}$ and $\varepsilon = 1/n$. Then by (91), for $k = 1, 2, \ldots$

$$\overline{p}[K(\overline{\zeta}_1, \ldots, \overline{\zeta}_k)] \geq \overline{p}[R(\overline{\zeta}_1, \ldots, \overline{\zeta}_k)] \geq \overline{p}(A) - (1/n) \geq c - (1/n),$$

so

$$\overline{p} \in \bigcap_{n=1}^{\infty} \bigcup_{z \in \mathscr{N}} \bigcap_{s < z} \{p \in P(X) | p[K(s)] \geq c - (1/n)\}.$$

On the other hand, given

$$\overline{p} \in \bigcap_{n=1}^{\infty} \bigcup_{z \in \mathscr{N}} \bigcap_{s < z} \{p \in P(X) | p[K(s)] \geq c - (1/n)\},$$

we have that for each $n$ there exists $(\zeta_1, \zeta_2, \ldots) \in \mathscr{N}$ for which

$$\overline{p}\left[\bigcap_{k=1}^{\infty} K(\zeta_1, \ldots, \zeta_k)\right] = \lim_{k \to \infty} \overline{p}[K(\zeta_1, \ldots, \zeta_k)] \geq c - (1/n).$$

We have then from (95) that

$$\overline{p}(A) \geq c - (1/n), \qquad n = 1, 2, \ldots,$$

so $\overline{p}(A) \geq c$, and (98) is proved.

Proposition 7.25 guarantees that for each $n \geq 1$ and $s \in \Sigma$, the set

$$T_n(s) = \{p \in P(X) | p[K(s)] \geq c - (1/n)\}$$

is Borel-measurable in $P(X)$. We have from (98) that

$$\{p \in P(X) | p(A) \geq c\} = \bigcap_{n=1}^{\infty} N(T_n),$$

and the proposition follows from Proposition 7.36 and Corollary 7.35.2.

Q.E.D.

**Corollary 7.43.1** Let $X$ be a Borel space and $A$ an analytic subset of $X$. For each $c \in R$, the set $\{p \in P(X) | p(A) > c\}$ is analytic.

*Proof* For each $c \in R$,

$$\{p \in P(X) | p(A) > c\} = \bigcup_{n=1}^{\infty} \{p \in P(X) | p(A) \geq c + (1/n)\}.$$

The result follows from Corollary 7.35.2 and Proposition 7.43. Q.E.D.

## 7.7 Lower Semianalytic Functions and Universally Measurable Selection

In a Borel space $X$, there are at least three $\sigma$-algebras which arise naturally†: the Borel $\sigma$-algebra $\mathscr{B}_X$ of Definition 7.6, the universal $\sigma$-algebra $\mathscr{U}_X$ of Definition 7.18, and the analytic $\sigma$-algebra $\mathscr{A}_X$, which we define now.

**Definition 7.19** Let $X$ be a Borel space. The *analytic $\sigma$-algebra* $\mathscr{A}_X$ is the smallest $\sigma$-algebra containing the analytic subsets of $X$. In symbols, $\mathscr{A}_X = \sigma[\mathscr{S}(\mathscr{F}_X)]$. If $E \in \mathscr{A}_X$, we say that $E$ is *analytically measurable.*

From Proposition 7.36 and Lusin's theorem (Proposition 7.42), we have that for any Borel space $X$

$$\mathscr{B}_X \subset \mathscr{S}(\mathscr{F}_X) \subset \mathscr{A}_X \subset \mathscr{U}_X.$$

If $X$ is countable, each of these collections of sets is equal to the power set of $X$ (the collection of all subsets of $X$). We show in Appendix B that if $X$ is uncountable, each set containment above is strict. This fact will not be used in the constructive part of the theory, but only to give examples showing that results cannot be strengthened.

Corresponding to the three $\sigma$-algebras just discussed, we will treat three types of measurability of functions. Borel-measurable functions were defined in Definition 7.8. The other two types are defined next.

**Definition 7.20** Let $X$ and $Y$ be Borel spaces and $f$ a function mapping $D \subset X$ into $Y$. If $D \in \mathscr{A}_X$ and $f^{-1}(B) \in \mathscr{A}_X$ for every $B \in \mathscr{B}_Y$, $f$ is said to be *analytically measurable.* If $D \in \mathscr{U}_X$ and $f^{-1}(B) \in \mathscr{U}_X$ for every $B \in \mathscr{B}_Y$, $f$ is said to be *universally measurable.*

From the preceding discussion, we see that every Borel-measurable function is analytically measurable, and every analytically measurable function is universally measurable. The converses of these statements are false.

We begin by stating for future reference the following characterization of the universal $\sigma$-algebra.

† A fourth $\sigma$-algebra, the *limit $\sigma$-algebra* $\mathscr{L}_X$, which lies between $\mathscr{A}_X$ and $\mathscr{U}_X$, is defined in Appendix B, and treated there and in Section 11.1.

**Lemma 7.26** Let $X$ be a Borel space and $E \subset X$. Then $E \in \mathscr{U}_X$ if and only if, given any $p \in P(X)$, there exists $B \in \mathscr{B}_X$ such that $p(E \triangle B) = 0$.

We turn now to the question of composition of measurable functions. If Borel-measurable functions are composed, the result is again Borel-measurable. Unfortunately, the composition of analytically measurable functions need not be analytically measurable (Appendix B). We have the following result for universally measurable functions.

**Proposition 7.44** Let $X$, $Y$, and $Z$ be Borel spaces, $D \in \mathscr{U}_X$, and $E \in \mathscr{U}_Y$. Suppose $f:D \to Y$ and $g:E \to Z$ are universally measurable and $f(D) \subset E$. Then the composition $g \circ f$ is universally measurable.

*Proof* We must show that given $B \in \mathscr{B}_Z$, the set $f^{-1}[g^{-1}(B)]$ is universally measurable. Since $g^{-1}(B) \in \mathscr{U}_Y$, it suffices to prove that $f^{-1}(U) \in \mathscr{U}_X$ for every $U \in \mathscr{U}_Y$. For $p \in P(X)$, define $p' \in P(Y)$ by

$$p'(C) = p[f^{-1}(C)] \qquad \forall C \in \mathscr{B}_Y.$$

Let $V \in \mathscr{B}_Y$ be such that

$$p[f^{-1}(V) \triangle f^{-1}(U)] = p'(V \triangle U) = 0.$$

The set $f^{-1}(V)$ is in $\mathscr{U}_X$, so there exists $W \in \mathscr{B}_X$ for which $p[W \triangle f^{-1}(V)] = 0$. Then $p[W \triangle f^{-1}(U)] = 0$. The result follows from Lemma 7.26. Q.E.D.

The proof of Proposition 7.44 also establishes the following fact.

**Corollary 7.44.1** Let $X$ and $Y$ be Borel spaces, $D \in \mathscr{U}_X$, and $f:D \to Y$ a universally measurable function. If $U \in \mathscr{U}_Y$, then $f^{-1}(U) \in \mathscr{U}_X$.

Since $\mathscr{A}_X \subset \mathscr{U}_X$, we can specialize these results to analytically measurable sets and functions.

**Corollary 7.44.2** Let $X$, $Y$, and $Z$ be Borel spaces, $D \in \mathscr{A}_X$, and $E \in \mathscr{A}_Y$. Suppose $f:D \to Y$ and $g:E \to Z$ are analytically measurable and $f(D) \subset E$. Then the composition $g \circ f$ is universally measurable. If $A \in \mathscr{A}_Y$, then $f^{-1}(A) \in \mathscr{U}_X$.

We remind the reader that if $X$ and $Y$ are Borel spaces, a stochastic kernel $q(dy|x)$ on $Y$ given $X$ is said to be universally measurable if the mapping $\gamma(x) = q(dy|x)$ is universally measurable from $X$ to $P(Y)$ (Definition 7.12).

**Corollary 7.44.3** Let $X$ and $Y$ be Borel spaces, let $f:X \to Y$ be a function, and let $q(dy|x)$ be a stochastic kernel on $Y$ given $X$ such that, for each $x$, $q(dy|x)$ assigns probability one to the point $f(x) \in Y$. Then $q(dy|x)$ is universally measurable if and only if $f$ is universally measurable.

*Proof* Let $\delta: Y \to P(Y)$ be the homeomorphism defined by $\delta(y) = p_y$ (Corollary 7.21.1). Let $\gamma: X \to P(Y)$ be the mapping $\gamma(x) = q(dy|x)$. Then $\gamma = \delta \circ f$ and $f = \delta^{-1} \circ \gamma$. The result follows from Proposition 7.44. Q.E.D.

If $X$ is a Borel space and $f: X \to R^*$ is universally measurable, then given any $p \in P(X)$, $f$ is measurable with respect to the completed Borel $\sigma$-algebra $\mathscr{B}_X(p)$, and $\int f\, dp$ is defined by

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

where the convention $\infty - \infty = \infty$ is used and the integrations are performed on the measure space $(X, \mathscr{B}_X(p), p)$. If $D \in \mathscr{U}_X$, the integral $\int_D f\, dp$ is defined similarly. Having thus defined $\int f\, dp$ without resort to $p$-outer measure, we have all the classical integration theorems at our disposal, provided that we take care with the addition of infinities.

We proceed now to show that universally measurable stochastic kernels can be used to define probability measures on product spaces in the manner of Proposition 7.28. For this we need some preparatory lemmas.

**Lemma 7.27** Let $X$ be a Borel space and $f: X \to R^*$. The function $f$ is universally measurable if and only if, for every $p \in P(X)$, there is a Borel-measurable function $f_p: X \to R^*$ such that $f(x) = f_p(x)$ for $p$ almost every $x$.

*Proof* Suppose $f$ is universally measurable and let $p \in P(X)$ be given. For $r \in Q^*$, let $U(r) = \{x | f(x) \le r\}$. Then $f(x) = \inf\{r \in Q^* | x \in U(r)\}$. Let $B(r) \in \mathscr{B}_X$ be such that $p[B(r) \triangle U(r)] = 0$. Define

$$f_p(x) = \inf\{r \in Q^* | x \in B(r)\} = \inf_{r \in Q^*} \psi_r(x),$$

where $\psi_r(x) = r$ if $x \in B(r)$ and $\psi_r(x) = \infty$ otherwise. Then $f_p: X \to R^*$ is Borel-measurable, and

$$\{x | f(x) \ne f_p(x)\} \subset \bigcup_{r \in Q^*} [B(r) \triangle U(r)]$$

has $p$-measure zero.

Conversely, if, given $p \in P(X)$, there is a Borel-measurable $f_p$ such that $f(x) = f_p(x)$ for $p$ almost every $x$, then

$$p(\{x | f(x) \le c\} \triangle \{x | f_p(x) \le c\}) = 0$$

for every $c \in R^*$, and the universal measurability of $f$ follows. Q.E.D.

Lemma 7.27 can be used to give an alternative definition of $\int f\, dp$ when $f$ is a universally measurable, extended real-valued function on a Borel space $X$ and $p \in P(X)$. Letting $f_p$ be as in the proof of that lemma, we can define $\int f\, dp = \int f_p\, dp$. It is easy to show that this definition is equivalent to the one which precedes Lemma 7.27.

**Lemma 7.28** Let $X$ and $Y$ be Borel spaces and let $q(dy|x)$ be a stochastic kernel on $Y$ given $X$. The following statements are equivalent:

(a) The stochastic kernel $q(dy|x)$ is universally measurable.
(b) For any $B \in \mathscr{B}_Y$, the mapping $\lambda_B: X \to R$ defined by $\lambda_B(x) = q(B|x)$ is universally measurable.
(c) For any $p \in P(X)$, there exists a Borel-measurable stochastic kernel $q_p(dy|x)$ on $Y$ given $X$ such that $q(dy|x) = q_p(dy|x)$ for $p$ almost every $x$.

*Proof* We show (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a). Assume (a) holds. Then the function $\gamma: X \to P(Y)$ given by $\gamma(x) = q(dy|x)$ is universally measurable. If $B \in \mathscr{B}_X$, $\lambda_B$ is defined as in (b), and $\theta_B: P(Y) \to R$ is given by $\theta_B(p) = p(B)$, then $\lambda_B = \theta_B \circ \gamma$, which is universally measurable by Propositions 7.25 and 7.44. Therefore (a) $\Rightarrow$ (b).

Assume (b) holds and choose $p \in P(X)$. Since $Y$ is separable and metrizable, there exists a countable base $\mathscr{B}$ for the topology in $Y$. Let $\mathscr{F}$ be the collection of sets in $\mathscr{B}$ and their finite intersections. For $F \in \mathscr{F}$, let $f_F$ be a Borel-measurable function for which

$$f_F(x) = q(F|x) \quad \forall x \in B_F,$$

where $B_F \in \mathscr{B}_X$ and $p(B_F) = 1$. Such an $f_F$ and $B_F$ exist by assumption (b) and Lemma 7.27. For $x \in \bigcap_{F \in \mathscr{F}} B_F$, let $q_p(dy|x) = q(dy|x)$. For $x \notin \bigcap_{F \in \mathscr{F}} B_F$, let $q_p(dy|x)$ be some fixed probability measure in $P(Y)$. Then $q(dy|x) = q_p(dy|x)$ for $p$ almost every $x$. The class of sets $\underline{Y}$ in $\mathscr{B}_Y$ for which $q_p(\underline{Y}|x)$ is Borel-measurable in $x$ is a Dynkin system containing $\mathscr{F}$. The class $\mathscr{F}$ is closed under finite intersections and generates $\mathscr{B}_Y$, so statement (c) follows from the Dynkin system theorem (Proposition 7.24). Therefore (b) $\Rightarrow$ (c).

Assume (c) holds and choose $p \in P(X)$. Let $q_p(dy|x)$ be as in assumption (c) and define $\gamma, \gamma_p: X \to P(Y)$ by $\gamma(x) = q(dy|x)$, $\gamma_p(x) = q_p(dy|x)$. If $B \in \mathscr{B}_{P(X)}$, then $p[\gamma^{-1}(B) \triangle \gamma_p^{-1}(B)] = 0$. Lemma 7.26 implies that $\gamma^{-1}(B)$ is universally measurable. Therefore (c) $\Rightarrow$ (a). Q.E.D.

**Lemma 7.29** Let $X$, $Y$, and $Z$ be Borel spaces and let $f: XY \to Z$ be a universally measurable function. For fixed $x \in X$, define $g_x: Y \to Z$ by

$$g_x(y) = f(x, y).$$

Then $g_x$ is universally measurable for every $x \in X$.

*Proof* For fixed $x_0 \in X$, let $\varphi: Y \to XY$ be the continuous function defined by $\varphi(y) = (x_0, y)$. For $\underline{Z} \in \mathscr{B}_Z$,

$$\{y \in Y | g_{x_0}(y) \in \underline{Z}\} = \varphi^{-1}(\{(x, y) \in XY | f(x, y) \in \underline{Z}\}),$$

and this set is universally measurable by Corollary 7.44.1. Q.E.D.

It is worth noting that if $(\Omega_1, \mathscr{F}_1, p)$ and $(\Omega_2, \mathscr{F}_2, q)$ are probability spaces, then there are two natural $\sigma$-algebras on $\Omega_1\Omega_2$, namely, $\mathscr{F}_1\mathscr{F}_2$ and the completion $\overline{\mathscr{F}_1\mathscr{F}_2}$ of $\mathscr{F}_1\mathscr{F}_2$ with respect to $pq$. If $f:\Omega_1\Omega_2 \to R$ is $\mathscr{F}_1\mathscr{F}_2$-measurable, then for every $\omega_1 \in \Omega_1$, the function $g_{\omega_1}(\omega_2) = f(\omega_1, \omega_2)$ is $\mathscr{F}_2$-measurable. However, if $f$ is only $\overline{\mathscr{F}_1\mathscr{F}_2}$-measurable, then $g_{\omega_1}(\omega_2)$ can be guaranteed to be $\mathscr{F}_2$-measurable only for $p$ almost all $\omega_1$. The case treated by Lemma 7.29 is intermediate to these two, since $\mathscr{U}_X\mathscr{U}_Y \subset \mathscr{U}_{XY}$, and if $p \in P(X)$, $q \in P(Y)$, and $\overline{\mathscr{U}_X\mathscr{U}_Y}$ denotes the completion of $\mathscr{U}_x\mathscr{U}_Y$ with respect to $pq$, then $\mathscr{U}_{XY} \subset \overline{\mathscr{U}_X\mathscr{U}_Y}$. Note that the stronger result that $g_x(y)$ is $\mathscr{U}_Y$-measurable for every $x \in X$ holds, although the assumption that $f$ is $\mathscr{U}_{XY}$-measurable may be weaker than the assumption that $f$ is $\mathscr{U}_X\mathscr{U}_Y$-measurable.

We now use the properties of universally measurable functions and stochastic kernels to extend Proposition 7.28.

**Proposition 7.45** Let $X_1, X_2, \ldots$ be a sequence of Borel spaces, $Y_n = X_1X_2\cdots X_n$ and $Y = X_1X_2\cdots$. Let $p \in P(X_1)$ be given and, for $n = 1, 2, \ldots$, let $q_n(dx_{n+1}|y_n)$ be a universally measurable stochastic kernel on $X_{n+1}$ given $Y_n$. Then for $n = 2, 3, \ldots$, there exist unique probability measures $r_n \in P(Y_n)$, such that

$$r_n(\underline{X}_1\underline{X}_2\cdots\underline{X}_n) = \int_{\underline{X}_1}\int_{\underline{X}_2}\cdots\int_{\underline{X}_{n-1}} q_{n-1}(\underline{X}_n|x_1, x_2, \ldots, x_{n-1}) \times q_{n-2}(dx_{n-1}|x_1, x_2, \ldots, x_{n-2})\cdots \times q_1(dx_2|x_1)p(dx_1) \qquad \forall \underline{X}_1 \in \mathscr{B}_{X_1}, \ldots, \underline{X}_n \in \mathscr{B}_{X_n}. \tag{99}$$

If $f: Y_n \to R^*$ is universally measurable and either $\int f^+\, dr_n < \infty$ or $\int f^-\, dr_n < \infty$, then

$$\int_{Y_n} f\, dr_n = \int_{X_1}\int_{X_2}\cdots\int_{X_n} f(x_1, x_2, \ldots, x_n)q_{n-1}(dx_n|x_1, x_2, \ldots, x_{n-1})\cdots \times q_1(dx_2|x_1)p(dx_1). \tag{100}$$

Furthermore, there exists a unique probability measure $r \in P(Y)$ such that for each $n$ the marginal of $r$ on $Y_n$ is $r_n$.

*Proof* There is a Borel-measurable stochastic kernel $\overline{q}_1(dx_2|x_1)$ which agrees with $q(dx_2|x_1)$ for $p$ almost every $x_1$. Define $r_2 \in P(Y_2)$ by specifying it on measurable rectangles to be (Proposition 7.28)

$$r_2(\underline{X}_1\underline{X}_2) = \int_{\underline{X}_1} \overline{q}_1(\underline{X}_2|x_1)p(dx_1) \qquad \forall \underline{X}_1 \in \mathscr{B}_{X_1}, \quad \underline{X}_2 \in \mathscr{B}_{X_2}.$$

Assume $f: Y_2 \to [0, \infty]$ is universally measurable and let $\overline{f}: Y_2 \to [0, \infty]$ be Borel-measurable and agree with $f$ on $Y_2 - N$, where $N \in \mathscr{B}_{Y_2}$ and $r_2(N) = 0$.

By Proposition 7.28,

$$0 = r_2(N) = \int_{X_1}\int_{X_2} \chi_N(x_1, x_2)\bar{q}_1(dx_2|x_1)p(dx_1)$$
$$= \int_{X_1} \bar{q}_1(N_{x_1}|x_1)p(dx_1),$$

so $\bar{q}_1(N_{x_1}|x_1) = 0$ for $p$ almost every $x_1$. Now $f(x_1, x_2) = \bar{f}(x_1, x_2)$ for $x_2 \notin N_{x_1}$ so

$$\left|\int_{X_2} [f(x_1, x_2) - \bar{f}(x_1, x_2)]\bar{q}_1(dx_2|x_1)\right|$$
$$\leq \int_{N_{x_1}} |f(x_1, x_2) - \bar{f}(x_1, x_2)|\bar{q}_1(dx_2|x_1) = 0$$

for $p$ almost every $x_1$. It follows that

$$\int_{X_1} \bar{f}(x_1, x_2)\bar{q}_1(dx_2|x_1) = \int_{X_2} f(x_1, x_2)\bar{q}_1(dx_2|x_1)$$
$$= \int_{X_2} f(x_1, x_2)q_1(dx_2|x_1)$$

for $p$ almost every $x_1$. The left-hand side is Borel-measurable by Proposition 7.29, so the right-hand side is universally measurable by Lemma 7.27. Furthermore,

$$\int_{Y_2} f\,dr_2 = \int_{Y_2} \bar{f}\,dr_2 = \int_{X_1}\int_{X_2} \bar{f}(x_1, x_2)\bar{q}_1(dx_2|x_1)p(dx_1)$$
$$= \int_{X_1}\int_{X_2} f(x_1, x_2)q_1(dx_2|x_1)p(dx_1).$$

This proves (100) for $n = 2$ and $f \geq 0$. If $f: Y_2 \to R^*$ is universally measurable and satisfies $\int f^+\,dr_2 < \infty$ or $\int f^-\,dr_2 < \infty$, then (100) holds for $f^+$ and $f^-$, so it holds for $f$ as well. Take $f = \chi_{\underline{X}_1\underline{X}_2}$ to obtain (99).

Now assume the proposition holds for $n = k$. Let $\bar{q}_k(dx_{k+1}|y_k)$ be a stochastic kernel which agrees with $q_k(dx_{k+1}|y_k)$ for $r_k$ almost every $x_k$. Define $r_{k+1}$ by specifying it on measurable rectangles to be

$$r_{k+1}(\underline{X}_1\underline{X}_2\cdots\underline{X}_{k+1}) = \int_{\underline{X}_1\underline{X}_2\cdots\underline{X}_k} \bar{q}_k(\underline{X}_{k+1}|x_1, x_2, \ldots, x_k)\,dr_k$$
$$\forall \underline{X}_1 \in \mathscr{B}_{X_1}, \ldots, \underline{X}_{k+1} \in \mathscr{B}_{X_{k+1}}.$$

Proceed as in the case of $n = 2$ to prove the proposition for $n = k + 1$. (See also the proof of Proposition 7.28.)

The existence of $r \in P(Y)$ such that the marginal of $r$ on $X_n$ is $r_n$, $n = 2, 3, \ldots$, is proved exactly as in Proposition 7.28. Q.E.D.

In the course of proving Proposition 7.45, we have also established the following fact.

**Proposition 7.46** Let $X$ and $Y$ be Borel spaces and let $f:XY \rightarrow R^*$ be universally measurable. Let $q(dy|x)$ be a universally measurable stochastic kernel on $Y$ given $X$. Then the mapping $\lambda:X \rightarrow R^*$ defined by

$$\lambda(x) = \int f(x, y)q(dy|x)$$

is universally measurable.

**Corollary 7.46.1** Let $X$ be a Borel space and let $f:X \rightarrow R^*$ be universally measurable. Then the function $\theta_f:P(X) \rightarrow R^*$ defined by

$$\theta_f(p) = \int f\, dp$$

is universally measurable.

*Proof* Define a universally measurable stochastic kernel on $X$ given $P(X)$ by $q(dx|p) = p(dx)$. Apply Proposition 7.46. Q.E.D.

As mentioned previously, the functions obtained by infimizing bivariate, extended real-valued, Borel-measurable functions over one of their variables have analytic lower level sets. We give these functions a name.

**Definition 7.21** Let $X$ be a Borel space, $D \subset X$, and $f:D \rightarrow R^*$. If $D$ is analytic and the set $\{x \in D | f(x) < c\}$ is analytic for every $c \in R$, then $f$ is said to be *lower semianalytic*.

It is apparent from the definition that a lower semianalytic function is analytically measurable. We state some characterizations and basic properties of lower semianalytic functions as a lemma.

**Lemma 7.30** (1) Let $X$ be a Borel space, $D$ an analytic subset of $X$, and $f:D \rightarrow X$. The following statements are equivalent.

(a) The function $f$ is lower semianalytic, i.e., the set

$$\{x \in D | f(x) < c\} \tag{101}$$

is analytic for every $c \in R$.

(b) The set (101) is analytic for every $c \in R^*$.

(c) The set

$$\{x \in D | f(x) \leq c\} \tag{102}$$

is analytic for every $c \in R$.

(d) The set (102) is analytic for every $c \in R^*$.

(2) Let $X$ be a Borel space, $D$ an analytic subset of $X$, and $f_n: D \to R^*$, $n = 1, 2, \ldots$, a sequence of lower semianalytic functions. Then the functions $\inf_n f_n$, $\sup_n f_n$, $\liminf_{n\to\infty} f_n$, and $\limsup_{n\to\infty} f_n$ are lower semianalytic. In particular, if $f_n \to f$, then $f$ is lower semianalytic.

(3) Let $X$ and $Y$ be Borel spaces, $g: X \to Y$, and $f: g(X) \to R^*$. If $g$ is Borel-measurable and $f$ is lower semianalytic, then $f \circ g$ is lower semianalytic.

(4) Let $X$ be a Borel space, $D$ an analytic subset of $X$, and $f, g: D \to R^*$. If $f$ and $g$ are lower semianalytic, then $f + g$ is lower semianalytic. If, in addition, $g$ is Borel-measurable and $g \geq 0$ or if $f \geq 0$ and $g \geq 0$, then $fg$ is lower semianalytic, where we define $0 \cdot \infty = \infty \cdot 0 = 0(-\infty) = (-\infty)0 = 0$.

*Proof* (1) We show (b) $\Rightarrow$ (a) $\Rightarrow$ (d) $\Rightarrow$ (c) $\Rightarrow$ (b). It is clear that (b) $\Rightarrow$ (a). If (a) holds, then

$$\{x \in D \mid f(x) \leq \infty\} = D$$

is analytic by definition, while the sets

$$\{x \in D \mid f(x) \leq -\infty\} = \bigcap_{n=1}^{\infty} \{x \in D \mid f(x) < -n\},$$

$$\{x \in D \mid f(x) \leq c\} = \bigcap_{n=1}^{\infty} \{x \in D \mid f(x) < c + (1/n)\}, \qquad c \in R,$$

are analytic by Corollary 7.35.2. Therefore (a) $\Rightarrow$ (d). It is clear that (d) $\Rightarrow$ (c). If (c) holds, then the sets

$$\{x \in D \mid f(x) < -\infty\} = \varnothing,$$

$$\{x \in D \mid f(x) < \infty\} = \bigcup_{n=1}^{\infty} \{x \in D \mid f(x) \leq n\},$$

$$\{x \in D \mid f(x) < c\} = \bigcup_{n=1}^{\infty} \{x \in D \mid f(x) \leq c - (1/n)\}, \qquad c \in R,$$

are analytic by Corollary 7.35.2. Therefore (c) $\Rightarrow$ (b).

(2) For $c \in R$,

$$\{x \in D \mid \inf_n f_n(x) < c\} = \bigcup_{n=1}^{\infty} \{x \in D \mid f_n(x) < c\},$$

$$\{x \in D \mid \sup_n f_n(x) \leq c\} = \bigcap_{n=1}^{\infty} \{x \in D \mid f_n(x) \leq c\},$$

so $\inf_n f_n$ and $\sup_n f_n$ are lower semianalytic by Corollary 7.35.2 and part (1). Then

$$\liminf_{n\to\infty} f_n = \sup_{n \geq 1} \inf_{k \geq n} f_k$$

and

$$\limsup_{n\to\infty} f_n = \inf_{n\geq 1} \sup_{k\geq n} f_k$$

are lower semianalytic as well.

(3) The domain $g(X)$ of $f$ is analytic by Proposition 7.40. For $c \in R$,

$$\{x \in X | (f \circ g)(x) < c\} = g^{-1}(\{y \in g(X) | f(y) < c\})$$

is analytic by the same proposition.

(4) For $c \in R$,

$$\{x \in D | f(x) + g(x) < c\} = \bigcup_{r\in Q} [\{x \in D | f(x) < r\} \cap \{x \in D | g(x) < c - r\}],$$

and this is true even if $f(x) + g(x) = \infty - \infty = \infty$ for some $x \in D$. From Corollary 7.35.2 it follows that $f + g$ is lower semianalytic whenever $f$ and $g$ are. Now suppose $g$ is Borel-measurable and $g \geq 0$. For $c > 0$, we have

$$\{x \in D | f(x)g(x) < c\} = \{x \in D | f(x) \leq 0\} \cup \{x \in D | g(x) \leq 0\} \cup \left[ \bigcup_{r\in Q, r>0} \{x \in D | f(x) < r, g(x) < c/r\} \right],$$

while if $c \leq 0$, we have

$$\{x \in D | f(x)g(x) < c\} = \bigcup_{r\in Q, r<0} \{x \in D | f(x) < r, g(x) > c/r\}.$$

In both cases, the set $\{x \in D | f(x)g(x) < c\}$ is analytic by Corollary 7.35.2. Suppose $f$ and $g$ are both lower semianalytic and nonnegative. For $c > 0$, the set $\{x \in D | f(x)g(x) < c\}$ is analytic as before, and for $c \leq 0$, this set is empty. It follows that $fg$ is lower semianalytic under either set of assumptions on $f$ and $g$. Q.E.D.

Note in connection with Lemma 7.30(3) that the composition of a Borel-measurable function with a lower semianalytic function can be guaranteed to be lower semianalytic only when the composition is in the order specified. To see this, let $X$ be a Borel space and $A \subset X$ be an analytic set whose complement is not analytic (see Appendix B). Define $f(x) = -\chi_A(x)$, which is lower semianalytic, because $\{x \in X | f(x) < c\}$ is either $\varnothing$, $A$, or $X$, depending on the value of $c$. Let $g: R^* \to R^*$ be given by $g(c) = -c$. Then $\chi_A = g \circ f$, and this function is not lower semianalytic, since $\{x \in X | \chi_A(x) < \frac{1}{2}\} = A^c$. This also provides us with an example of an analytically measurable function which is not lower semianalytic.

**Proposition 7.47** Let $X$ and $Y$ be Borel spaces, let $D$ be an analytic subset of $XY$, and let $f: D \to R^*$ be lower semianalytic. Then the function

$f^*: \text{proj}_X(D) \to R^*$ defined by

$$f^*(x) = \inf_{y \in D_x} f(x, y) \tag{103}$$

is lower semianalytic. Conversely, if $f^*: X \to R^*$ is a given lower semianalytic function and $Y$ is an uncountable Borel space, then there exists a Borel-measurable function $f: XY \to R^*$ which satisfies (103) with $D = XY$.

*Proof* For the first part of the theorem, observe that if $f: D \to R^*$ is lower semianalytic and $c \in R$, the set

$$\left\{x \in \text{proj}_X(D) \middle| \inf_{y \in D_x} f(x, y) < c\right\} = \text{proj}_X(\{(x, y) \in D | f(x, y) < c\})$$

is analytic by Proposition 7.39.

For the converse part of the theorem, let $f^*: X \to R^*$ be lower semianalytic and let $Y$ be an uncountable Borel space. For $r \in Q$, let $A(r) = \{x \in X | f^*(x) < r\}$. Then $A(r)$ is analytic and, by Proposition 7.39, there exists $B(r) \in \mathscr{B}_{XY}$ such that $A(r) = \text{proj}_X[B(r)]$. Define $G(r) = \bigcup_{s \in Q, s \leq r} B(s)$ and $f: XY \to R^*$ by

$$f(x, y) = \inf\{r \in Q | (x, y) \in G(r)\} = \inf_{r \in Q} \psi_r(x, y),$$

where $\psi_r(x, y) = r$ if $(x, y) \in G(r)$ and $\psi_r(x, y) = \infty$ otherwise. Then $f$ is Borel-measurable. Let $g$ be defined by $g(x) = \inf_{y \in Y} f(x, y)$. We show that $f^*(x) = g(x)$ for every $x \in X$.

If $f^*(x) < c$ for some $c \in R$, then there exists $r \in Q$ for which $f^*(x) < r < c$, and so $x \in A(r)$. There exists $y \in Y$ such that $(x, y) \in G(r)$, and, consequently, $f(x, y) \leq r$ and $g(x) \leq r < c$. Therefore $g(x)$ cannot be greater than $f^*(x)$.

If $g(x) < c$ for some $c \in R$, then there exists $r \in Q$ and $y \in Y$ for which $g(x) < r < c$ and $(x, y) \in G(r)$. Thus for some $s \in Q$, $s \leq r$, we have $(x, y) \in B(s)$ and $x \in A(s)$. This implies $f^*(x) < s \leq r < c$, which shows that $f^*(x)$ cannot be greater than $g(x)$. Q.E.D.

**Proposition 7.48** Let $X$ and $Y$ be Borel spaces, $f: XY \to R^*$ lower semianalytic, and $q(dy|x)$ a Borel-measurable stochastic kernel on $Y$ given $X$. Then the function $\lambda: X \to R^*$ defined by

$$\lambda(x) = \int f(x, y) q(dy|x)$$

is lower semianalytic.

*Proof* Suppose $f \geq 0$. Let $f_n(x, y) = \min\{n, f(x, y)\}$. Then each $f_n$ is lower semianalytic and $f_n \uparrow f$. The set

$$\begin{aligned} E_n &= \{(x, y, b) \in XYR | f_n(x, y) \leq b \leq n\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{r \in Q} \{(x, y, b) \in XYR | f_n(x, y) < r, r \leq b + (1/k) \leq n + (1/k)\} \end{aligned}$$

is analytic in $XYR$ by Corollary 7.35.2 and Proposition 7.38. Let $\mu$ be Lebesgue measure on $R$, $p \in P(XY)$, and $p\mu$ the product measure on $XYR$. By Fubini's theorem,

$$(p\mu)(E_n) = \int_{XY} \int_R \chi_{E_n} \, d\mu \, dp = \int_{XY} [n - f_n(x, y)] \, dp$$
$$= n - \int_{XY} f_n(x, y) \, dp.$$

For $c \in R$ we have, by the monotone convergence theorem,

$$\left\{p \in P(XY) \middle| \int f(x, y) \, dp \le c\right\} = \bigcap_{n=1}^{\infty} \left\{p \in P(XY) \middle| \int_{XY} f_n(x, y) \, dp \le c\right\}$$
$$= \bigcap_{n=1}^{\infty} \{p \in P(XY) | (p\mu)(E_n) \ge n - c\}.$$

Hence, by Proposition 7.43 and the fact that the mapping $p \to p\mu$ is continuous (Lemma 7.12), the function $\theta_f : P(XY) \to R^*$ defined by $\theta_f(p) = \int f(x, y) \, dp$ is lower semianalytic. We have

$$\lambda(x) = \theta_f[q(dy|x)p_x].$$

Since the mapping $x \to q(dy|x)$ is Borel-measurable from $X$ to $P(Y)$ and the mappings $x \to p_x$ and $[q(dy|x), p_x] \to q(dy|x)p_x$ are continuous from $X$ to $P(X)$ and $P(X)P(Y)$ to $P(XY)$, respectively (Corollary 7.21.1 and Lemma 7.12), it follows from Lemma 7.30(3) that $\lambda$ is lower semianalytic.

Suppose $f \le 0$. Let $f_n(x, y) = \max\{-n, f(x, y)\}$. Then each $f_n$ is lower semianalytic and $f_n \downarrow f$. The sets $E_n = \{(x, y, b) \in XYR | f_n(x, y) \le b \le 0\}$ are analytic and

$$(p\mu)(E_n) = \int_{XY} \int_R \chi_{E_n} \, d\mu \, dp = -\int_{XY} f_n(x, y) \, dp.$$

For $c \in R$,

$$\left\{p \in P(XY) \middle| \int f(x, y) \, dp < c\right\} = \bigcup_{n=1}^{\infty} \left\{p \in P(XY) \middle| \int_{XY} f_n(x, y) \, dp < c\right\}$$
$$= \bigcup_{n=1}^{\infty} \{p \in P(XY) | (p\mu)(E_n) > -c\}.$$

Proceed as before.

In the general case,

$$\int f(x, y) q(dy|x) = \int f^+(x, y) q(dy|x) - \int f^-(x, y) q(dy|x).$$

The functions $f^+$ and $-f^-$ are lower semianalytic, so by the preceding arguments each of the summands on the right is lower semianalytic. The result follows from Lemma 7.30(4). Q.E.D.

**Corollary 7.48.1** Let $X$ be a Borel space and let $f: X \to R^*$ be lower semianalytic. Then the function $\theta_f: P(X) \to R^*$ defined by

$$\theta_f(p) = \int f \, dp$$

is lower semianalytic.

*Proof* Define a Borel-measurable stochastic kernel on $X$ given $P(X)$ by $q(dx|p) = p(dx)$. Apply Proposition 7.48. Q.E.D.

As an aid in proving the selection theorem for lower semianalytic functions, we give a result concerning selection in an analytic subset of a product of Borel spaces. The reader will notice a strong resemblance between this result and Lemma 7.21, which was instrumental in proving the selection theorem for upper semicontinuous functions.

**Proposition 7.49** (Jankov–von Neumann theorem) Let $X$ and $Y$ be Borel spaces and $A$ an analytic subset of $XY$. There exists an analytically measurable function $\varphi: \text{proj}_X(A) \to Y$ such that $\text{Gr}(\varphi) \subset A$.

*Proof* (See Fig. 7.2.) Let $f: \mathcal{N} \to XY$ be continuous such that $A = f(\mathcal{N})$. Let $g = \text{proj}_X \circ f$. Then $g: \mathcal{N} \to X$ is continuous from $\mathcal{N}$ onto $\text{proj}_X(A)$. For $x \in \text{proj}_X(A)$, $g^{-1}(\{x\})$ is a closed nonempty subset of $\mathcal{N}$. Let $\zeta_1(x)$ be the smallest integer which is the first component of an element $z_1 \in g^{-1}(\{x\})$. Let $\zeta_2(x)$ be the smallest integer which is the second component of an element $z_2 \in g^{-1}(\{x\})$ whose first component is $\zeta_1(x)$. In general, let $\zeta_k(x)$ be the smallest integer which is the $k$th component of an element $z_k \in g^{-1}(\{x\})$ whose first $(k-1)$st components are $\zeta_1(x), \ldots, \zeta_{k-1}(x)$. Let $\psi(x) = (\zeta_1(x), \zeta_2(x), \ldots)$. Since $z_k \to \psi(x)$, we have

$$\psi(x) \in g^{-1}(\{x\}). \tag{104}$$

Define $\varphi: \text{proj}_X(A) \to Y$ by $\varphi = \text{proj}_Y \circ f \circ \psi$, so that $\text{Gr}(\varphi) \subset A$.

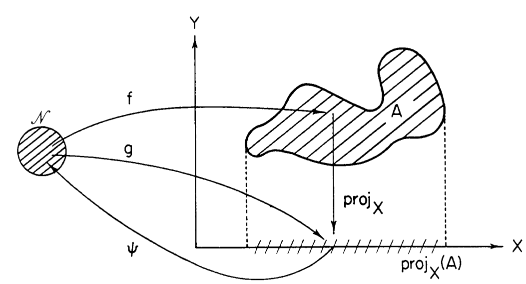


FIGURE 7.2

We show that $\varphi$ is analytically measurable. As in the proof of Proposition 7.42, for $(\sigma_1,\ldots,\sigma_k)\in\Sigma$ let

$$N(\sigma_1,\ldots,\sigma_k) = \{(\zeta_1,\zeta_2,\ldots)\in\mathscr{N}|\zeta_1 = \sigma_1,\ldots,\zeta_k = \sigma_k\},$$
$$M(\sigma_1,\ldots,\sigma_k) = \{(\zeta_1,\zeta_2,\ldots)\in\mathscr{N}|\zeta_1 \le \sigma_1,\ldots,\zeta_k \le \sigma_k\}.$$

We first show that $\psi$ is analytically measurable, i.e., $\psi^{-1}(\mathscr{B}_{\mathscr{N}}) \subset \mathscr{A}_X$. Since $\{N(s)|s\in\Sigma\}$ is a base for the topology on $\mathscr{N}$, by the remark following Definition 7.6, we have $\sigma(\{N(s)|s\in\Sigma\}) = \mathscr{B}_{\mathscr{N}}$. Then

$$\psi^{-1}(\mathscr{B}_{\mathscr{N}}) = \psi^{-1}[\sigma(\{N(s)|s\in\Sigma\})] = \sigma[\psi^{-1}(\{N(s)|s\in\Sigma\})],$$

and it suffices to prove

$$\psi^{-1}[N(s)]\in\mathscr{A}_X \qquad \forall s\in\Sigma. \tag{105}$$

We claim that for $s = (\sigma_1,\sigma_2,\ldots,\sigma_k)\in\Sigma$

$$\psi^{-1}[N(s)] = g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j - 1)], \tag{106}$$

where $M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j - 1) = \varnothing$ if $\sigma_j - 1 = 0$. We show this by proving that $\psi^{-1}[N(s)]$ is a subset of the set on the right-hand side of (106) and vice versa.

Suppose $x\in\psi^{-1}[N(s)]$. Let $\psi(x) = (\zeta_1(x),\zeta_2(x),\ldots)$. Then

$$\psi(x)\in N(s) \subset M(s), \tag{107}$$

so (104) implies

$$x = g[\psi(x)]\in g[M(s)]. \tag{108}$$

Relation (107) also implies $\zeta_1(x) = \sigma_1,\ldots,\zeta_k(x) = \sigma_k$. By the construction of $\psi$, we have that $\sigma_1$ is the smallest integer which is the first component of an element of $g^{-1}(\{x\})$, and for $j = 2,\ldots,k$, $\sigma_j$ is the smallest integer which is the $j$th component of an element of $g^{-1}(\{x\})$ whose first $(j-1)$ components are $\sigma_1,\ldots,\sigma_{j-1}$. In other words,

$$g^{-1}(\{x\}) \cap M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j - 1) = \varnothing, \qquad j = 1,\ldots,k.$$

It follows that

$$x\notin \bigcup_{j=1}^{k} g[M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j - 1)]. \tag{109}$$

Relations (108) and (109) imply

$$\psi^{-1}[N(s)] \subset g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j - 1)]. \tag{110}$$

To prove the reverse set containment, suppose

$$x \in g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1, \ldots, \sigma_{j-1}, \sigma_j - 1)]. \tag{111}$$

Since $x \in g[M(s)]$, there must exist $y = (\eta_1, \eta_2, \ldots) \in g^{-1}(\{x\})$ such that

$$\eta_1 \le \sigma_1, \ldots, \eta_k \le \sigma_k. \tag{112}$$

Clearly, $x \in \mathrm{proj}_X(A) = g(\mathcal{N})$, so $\psi(x)$ is defined. Let $\psi(x) = (\zeta_1(x), \zeta_2(x), \ldots)$. By (104), we have $g[\psi(x)] = x$, so (111) implies

$$\psi(x) \notin M(\sigma_1, \ldots, \sigma_{j-1}, \sigma_j - 1), \qquad j = 1, 2, \ldots, k.$$

Since $\psi(x) \notin M(\sigma_1 - 1)$, we know that $\zeta_1(x) \ge \sigma_1$. But $\zeta_1(x)$ is the smallest integer which is the first component of an element of $g^{-1}(\{x\})$, so (112) implies $\zeta_1(x) \le \eta_1 \le \sigma_1$. Therefore $\zeta_1(x) = \sigma_1$. Similarly, since $\psi(x) \notin M(\zeta_1(x), \sigma_2 - 1)$, we have $\zeta_2(x) \ge \sigma_2$, Again from (112) we see that $\zeta_2(x) \le \eta_2 \le \sigma_2$, so $\zeta_2(x) = \sigma_2$. Continuing in this manner, we show that $\psi(x) \in N(s)$, i.e., $x \in \psi^{-1}[N(s)]$ and

$$\psi^{-1}[N(s)] \supset g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1, \ldots, \sigma_{j-1}, \sigma_j - 1)]. \tag{113}$$

Relations (110) and (113) imply (106).

We note now that $M(t)$ is open in $\mathcal{N}$ for every $t \in \Sigma$, so $g[M(t)]$ is analytic by Proposition 7.40. Relation (105) now follows from (106), so $\psi$ is analytically measurable.

By the definition of $\varphi$ and the Borel-measurability of $f$ and $\mathrm{proj}_Y$, we have

$$\varphi^{-1}(\mathscr{B}_Y) = \psi^{-1}(f^{-1}[\mathrm{proj}_Y^{-1}(\mathscr{B}_Y)]) \subset \psi^{-1}(f^{-1}[\mathscr{B}_{XY}]) \subset \psi^{-1}(\mathscr{B}_{\mathcal{N}}).$$

We have just proved $\psi^{-1}(\mathscr{B}_{\mathcal{N}}) \subset \mathscr{A}_X$, and the analytic measurability of $\varphi$ follows. Q.E.D.

This brings us to the selection theorem for lower semianalytic functions.

**Proposition 7.50** Let $X$ and $Y$ be Borel spaces, $D \subset XY$ an analytic set, and $f: D \to R^*$ a lower semianalytic function. Define $f^*: \mathrm{proj}_X(D) \to R^*$ by

$$f^*(x) = \inf_{y \in D_x} f(x, y). \tag{114}$$

(a) For every $\varepsilon > 0$, there exists an analytically measurable function $\varphi: \mathrm{proj}_X(D) \to Y$ such that $\mathrm{Gr}(\varphi) \subset D$ and for all $x \in \mathrm{proj}_X(D)$,

$$f[x, \varphi(x)] \le \begin{cases} f^*(x) + \varepsilon & \text{if } f^*(x) > -\infty, \\ -1/\varepsilon & \text{if } f^*(x) = -\infty. \end{cases}$$

(b) The set

$$I = \{x \in \operatorname{proj}_X(D) | \text{for some } y_x \in D_x, f(x, y_x) = f^*(x)\}$$

is universally measurable, and for every $\varepsilon > 0$ there exists a universally measurable function $\varphi : \operatorname{proj}_X(D) \to Y$ such that $\operatorname{Gr}(\varphi) \subset D$ and for all $x \in \operatorname{proj}_X(D)$

$$f[x, \varphi(x)] = f^*(x) \qquad \text{if} \quad x \in I, \tag{115}$$

$$f[x, \varphi(x)] \leq \begin{cases} f^*(x) + \varepsilon & \text{if} \quad x \notin I, \quad f^*(x) > -\infty, \\ -1/\varepsilon & \text{if} \quad x \notin I, \quad f^*(x) = -\infty. \end{cases} \tag{116}$$

*Proof* (a) (Cf. proof of Proposition 7.34 and Fig. 7.1.) The function $f^*$ is lower semianalytic by Proposition 7.47. For $k = 0, \pm 1, \pm 2, \ldots$, define

$$\begin{aligned} A(k) &= \{(x, y) \in D | f(x, y) < k\varepsilon\}, \\ B(k) &= \{x \in \operatorname{proj}_X(D) | (k-1)\varepsilon \leq f^*(x) < k\varepsilon\}, \\ B(-\infty) &= \{x \in \operatorname{proj}_X(D) | f^*(x) = -\infty\}. \\ B(\infty) &= \{x \in \operatorname{proj}_X(D) | f^*(x) = \infty\}. \end{aligned}$$

The sets $A(k)$, $k = 0, \pm 1, \pm 2, \ldots$, and $B(-\infty)$ are analytic, while the sets $B(k)$, $k = 0, \pm 1, \pm 2, \ldots$, and $B(\infty)$ are analytically measurable. By the Jankov–von Neumann theorem (Proposition 7.49) there exists, for each $k = 0, \pm 1, \pm 2, \ldots$, an analytically measurable $\varphi_k : \operatorname{proj}_X[A(k)] \to Y$ with $(x, \varphi_k(x)) \in A(k)$ for all $x \in \operatorname{proj}_X[A(k)]$ and an analytically measurable $\bar{\varphi} : \operatorname{proj}_X(D) \to Y$ such that $(x, \bar{\varphi}(x)) \in D$ for all $x \in \operatorname{proj}_X(D)$. Let $k^*$ be an integer such that $k^* \leq -1/\varepsilon^2$. Define $\varphi : \operatorname{proj}_X(D) \to Y$ by

$$\varphi(x) = \begin{cases} \varphi_k(x) & \text{if} \quad x \in B(k), \quad k = 0, \pm 1, \pm 2, \ldots, \\ \bar{\varphi}(x) & \text{if} \quad x \in B(\infty), \\ \varphi_{k^*}(x) & \text{if} \quad x \in B(-\infty). \end{cases}$$

Since $B(k) \subset \operatorname{proj}_X[A(k)]$ and $B(-\infty) \subset \operatorname{proj}_X[A(k)]$ for all $k$, this definition is possible. It is clear that $\varphi$ is analytically measurable and $\operatorname{Gr}(\varphi) \subset D$. If $x \in B(k)$, then $(x, \varphi_k(x)) \in A(k)$ and we have

$$f[x, \varphi(x)] = f[x, \varphi_k(x)] < k\varepsilon \leq f^*(x) + \varepsilon.$$

If $x \in B(\infty)$, then $f(x, y) = \infty$ for all $y \in D_x$ and $f[x, \varphi(x)] = \infty = f^*(x)$. If $x \in B(-\infty)$, we have

$$f[x, \varphi(x)] = f[x, \varphi_{k^*}(x)] < k^*\varepsilon \leq -1/\varepsilon.$$

Hence $\varphi$ has the required properties.

(b) Consider the set $E \subset XYR^*$ defined by

$$E = \{(x, y, b) | (x, y) \in D, f(x, y) \leq b\}.$$

Since

$$E = \bigcap_{k=1}^{\infty} \bigcup_{r \in Q^*} \{(x, y, b) | (x, y) \in D, f(x, y) \leq r, r \leq b + (1/k)\},$$

it follows from Corollary 7.35.2 and Proposition 7.38 that $E$ is analytic in $XYR^*$, and hence the set

$$A = \text{proj}_{XR^*}(E)$$

is analytic in $XR^*$. The mapping $T: \text{proj}_X(D) \to XR^*$ defined by

$$T(x) = (x, f^*(x))$$

is analytically measurable, and

$$I = \{x | (x, f^*(x)) \in A\} = T^{-1}(A).$$

Hence $I$ is universally measurable by Corollary 7.44.2.

Since $E$ is analytic, there is, by the Jankov–von Neumann Theorem, an analytically measurable $\rho: A \to Y$ such that $(x, \rho(x, b), b) \in E$ for every $(x, b) \in A$. Define $\psi: I \to Y$ by

$$\psi(x) = \rho(x, f^*(x)) = (\rho \circ T)(x) \qquad \forall x \in I.$$

Then $\psi$ is universally measurable by Corollary 7.44.2, and by construction $f[x, \psi(x)] \leq f^*(x)$ for $x \in I$. Hence

$$f[x, \psi(x)] = f^*(x) \qquad \forall x \in I. \tag{117}$$

By part (a) there exists an analytically measurable $\psi_\varepsilon: \text{proj}_X(D) \to Y$ such that

$$f[x, \psi_\varepsilon(x)] \leq \begin{cases} f^*(x) + \varepsilon & \text{if } f^*(x) > -\infty, \\ -1/\varepsilon & \text{if } f^*(x) = -\infty. \end{cases} \tag{118}$$

Define $\varphi: \text{proj}_X(D) \to Y$ by

$$\varphi(x) = \begin{cases} \psi(x) & \text{if } x \in I, \\ \psi_\varepsilon(x) & \text{if } x \in \text{proj}_X(D) - I. \end{cases}$$

Then $\varphi$ is universally measurable and, by (117) and (118), it has the required properties. Q.E.D.

Since the composition of analytically measurable functions can fail to be analytically measurable (Appendix B), the selector obtained in the proof of Proposition 7.50(b) can fail to be analytically measurable. The composition of universally measurable functions is universally measurable, and so we obtained a selector which is universally measurable. However, there is a $\sigma$-algebra, which we call the *limit $\sigma$-algebra*, lying between $\mathscr{A}_X$ and $\mathscr{U}_X$ such that the composition of limit measurable functions is again limit-measurable. We discuss this $\sigma$-algebra in Appendix B and state a strengthened version of Proposition 7.50 in Section 11.1.