Chapter 8

The Finite Horizon Borel Model

In Chapters 8–10 we will treat a model very similar to that of Section 2.3.2. An applications-oriented treatment of that model can be found in "Dynamic Programming and Stochastic Control" by Bertsekas [B4], hereafter referred to as DPSC. The model of Section 2.3.2 and DPSC has a countable disturbance space and arbitrary state and control spaces, whereas the model treated here will have Borel state, control, and disturbance spaces.

8.1 The Model

Definition 8.1 A *finite horizon stochastic optimal control model* is a nine-tuple $(S, C, U, W, p, f, \alpha, g, N)$ as described here. The letters x and u are used to denote elements of S and C, respectively.

S *State space.* A nonempty Borel space.
C *Control space.* A nonempty Borel space.
U *Control constraint.* A function from S to the set of nonempty subsets of C. The set

$$\Gamma = \{(x, u) | x \in S, u \in U(x)\} \tag{1}$$

is assumed to be analytic in SC.

W *Disturbance space.* A nonempty Borel space.

$p(dw|x,u)$ *Disturbance kernel.* A Borel-measurable stochastic kernel on W given SC.

f *System function.* A Borel-measurable function from SCW to S.

α *Discount factor.* A positive real number.

g *One-stage cost function.* A lower semianalytic function from Γ to R^*.

N *Horizon.* A positive integer.

We envision a system moving from state x_k to state x_{k+1} via the system equation

$$x_{k+1} = f(x_k, u_k, w_k), \qquad k = 0, 1, \ldots, N-2,$$

and incurring cost at each stage of $g(x_k, u_k)$. The disturbances w_k are random objects with probability distributions $p(dw_k|x_k, u_k)$. The goal is to choose u_k dependent on the history $(x_0, u_0, \ldots, x_{k-1}, u_{k-1}, x_k)$ so as to minimize

$$E\left\{\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k)\right\}. \tag{2}$$

The meaning of this statement will be made precise shortly. We have the constraint that when x_k is the kth state, the kth control u_k must be chosen to lie in $U(x_k)$.

In the models in Section 2.3.2 and DPSC, the one-stage cost g is also a function of the disturbance, i.e., has the form $g(x, u, w)$. If this is the case, then $g(x, u, w)$ can be replaced by

$$\bar{g}(x, u) = \int g(x, u, w) p(dw|x, u).$$

If $g(x, u, w)$ is lower semianalytic, so is $\bar{g}(x, u)$ (Proposition 7.48). If $p(dw|x, u)$ is continuous and $g(x, u, w)$ is lower semicontinuous and bounded below or upper semicontinuous and bounded above, then $\bar{g}(x, u)$ is lower semicontinuous and bounded below or upper semicontinuous and bounded above, respectively (Proposition 7.31). Since these are the three cases we deal with, there is no loss of generality in considering a one-stage cost function which is independent of the disturbance.

The model posed in Definition 8.1 is stationary, i.e., the data does not vary from stage to stage. A reduction of the nonstationary model to this form is discussed in Section 10.1.

A notational device which simplifies the presentation is the *state transition stochastic kernel* on S given SC defined by

$$t(B|x, u) = p(\{w|f(x, u, w) \in B\}|x, u) = p(f^{-1}(B)_{(x,u)}|x, u). \tag{3}$$

Thus $t(B|x,u)$ is the probability that the $(k+1)$st state is in B given that the kth state is x and the kth control is u. Proposition 7.26 and Corollary 7.26.1 imply that $t(dx'|x,u)$ is Borel-measurable.

Definition 8.2 A *policy* for the model of Definition 8.1 is a sequence $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ such that, for each k, $\mu_k(du_k|x_0, u_0, \ldots, u_{k-1}, x_k)$ is a universally measurable stochastic kernel on C given $SC\cdots CS$ satisfying

$$\mu_k(U(x_k)|x_0, u_0, \ldots, u_{k-1}, x_k) = 1$$

for every $(x_0, u_0, \ldots, u_{k-1}, x_k)$. If, for each k, μ_k is parameterized only by (x_0, x_k), π is a *semi-Markov policy*. If μ_k is parameterized only by x_k, π is a *Markov policy*. If, for each k and $(x_0, u_0, \ldots, u_{k-1}, x_k)$, $\mu_k(du_k|x_0, u_0, \ldots, u_{k-1}, x_k)$ assigns mass one to some point in C, π is *nonrandomized*. In this case, by a slight abuse of notation, π can be considered to be a sequence of universally measurable (Corollary 7.44.3) mappings $\mu_k: SC\cdots CS \to C$ such that

$$\mu_k(x_0, u_0, \ldots, u_{k-1}, x_k) \in U(x_k)$$

for every $(x_0, u_0, \ldots, u_{k-1}, x_k)$. If $\mathscr{F}$ is a type of σ-algebra on Borel spaces and all the stochastic kernel components of a policy are $\mathscr{F}$-measurable, we say the *policy is $\mathscr{F}$-measurable*. (For example, $\mathscr{F}$ could represent the Borel σ-algebras or the analytic σ-algebras.)

We denote by Π' *the set of all policies for the model of Definition 8.1 and by* Π *the set of all Markov policies.* We will show that in many cases it is not necessary to go outside Π to find the "best" available policy. In most cases, this "best" policy can be taken to be nonrandomized. Since Γ is analytic, the Jankov–von Neumann theorem (Proposition 7.49) guarantees that there exists at least one nonrandomized Markov policy, so Π and Π' are nonempty.

If $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ is a nonrandomized Markov policy, then π is a finite horizon version of a policy in the sense of Section 2.1. The notion of policy as set forth in Definition 8.2 is wider than the concept of Section 2.1 in that randomized non-Markov policies are permitted. It is narrower in that universal measurability is required.

We are now in a position to make precise expression (2). In this and subsequent discussions, we often index the state and control spaces for clarity. However, except in Chapter 10 when the nonstationary model is treated, we will always understand S_k to be a copy of S and C_k to be a copy of C. Suppose $p \in P(S)$ and $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ is a policy for the model of Definition 8.1. By Proposition 7.45, there is a unique probability measure $r_N(\pi, p)$ on $S_0C_0\cdots S_{N-1}C_{N-1}$ such that for any universally measurable function $h: S_0C_0\cdots S_{N-1}C_{N-1} \to R^*$ which satisfies either $\int h^+ \, dr_N(\pi, p) < \infty$

or $\int h^- \, dr_N(\pi, p) < \infty$, we have

$$\begin{aligned}\int h \, dr_N(\pi, p) = \int_{S_0}\int_{C_0}\int_{S_1}\cdots\int_{S_{N-1}}\int_{C_{N-1}} & h(x_0, u_0, \ldots, x_{N-1}, u_{N-1}) \\ & \times \mu_{N-1}(du_{N-1}|x_0, u_0, \ldots, u_{N-2}, x_{N-1}) \\ & \times t(dx_{N-1}|x_{N-2}, u_{N-2})\cdots t(dx_1|x_0, u_0)\mu_0(du_0|x_0)p(dx_0), \quad (4)\end{aligned}$$

where $t(dx'|x, u)$ is the Borel-measurable stochastic kernel defined by (3). Furthermore we have from (4) that $\int h dr_N(\pi, p_x)$ is a universally measurable function of x (Proposition 7.46), and if h and π are Borel-measurable, then $\int h dr_N(\pi, p_x)$ is a Borel-measurable function of x (Proposition 7.29).

Definition 8.3 Suppose $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ is a policy for the model of Definition 8.1. For $K \leq N$, the *K-stage cost corresponding to* π at $x \in S$ is

$$J_{K,\pi}(x) = \int \left[\sum_{k=0}^{K-1} \alpha^k g(x_k, u_k)\right] dr_N(\pi, p_x), \quad (5)$$

where, for each $\pi \in \Pi'$ and $p \in P(S)$, $r_N(\pi, p)$ is the unique probability measure satisfying (4). The *K-stage optimal cost* at x is

$$J_K^*(x) = \inf_{\pi \in \Pi'} J_{K,\pi}(x). \quad (6)$$

If $\varepsilon > 0$, the policy π is *K-stage ε-optimal at* x provided

$$J_{K,\pi}(x) \leq \begin{cases} J_K^*(x) + \varepsilon & \text{if } J_K^*(x) > -\infty, \\ -1/\varepsilon & \text{if } J_K^*(x) = -\infty. \end{cases}$$

If $J_{K,\pi}(x) = J_K^*(x)$, then π is *K-stage optimal at* x. If π is K-stage ε-optimal or K-stage optimal at every $x \in S$, it is said to be *K-stage ε-optimal* or *K-stage optimal*, respectively. If $\{\varepsilon_n\}$ is a sequence of positive numbers with $\varepsilon_n \downarrow 0$, a sequence of policies $\{\pi_n\}$ exhibits *$\{\varepsilon_n\}$-dominated convergence to K-stage optimality* provided

$$\lim_{n\to\infty} J_{K,\pi_n} = J_K^*,$$

and for $n = 2, 3, \ldots$

$$J_{K,\pi_n}(x) \leq \begin{cases} J_K^*(x) + \varepsilon_n & \text{if } J_K^*(x) > -\infty, \\ J_{K,\pi_{n-1}}(x) + \varepsilon_n & \text{if } J_K^*(x) = -\infty. \end{cases}$$

If $K = N$, we suppress the qualifier "K-stage" in the preceding terms.

Note that J_K^* is independent of the horizon N as long as $K \leq N$. Note also that $J_{K,\pi}(x)$ is universally measurable in x. If π is a Borel-measurable policy and g is Borel-measurable, then $J_{K,\pi}(x)$ is Borel-measurable in x.

For $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1}) \in \Pi'$ and $p \in P(S)$, let $q_k(\pi, p)$ be the marginal of $r_N(\pi, p)$ on $S_k C_k$. If we take $h = \chi_{S_0 \cdots C_{k-1} \underline{S}_k \underline{C}_k S_{k+1} \cdots C_{N-1}}$ in (4), we obtain

$$\begin{aligned} q_k(\pi, p)(\underline{S}_k \underline{C}_k) &= \int_{S_0} \int_{C_0} \int_{S_1} \cdots \int_{C_{k-1}} \int_{\underline{S}_k} \mu_k(\underline{C}_k | x_0, u_0, \ldots, u_{k-1}, x_k) \\ &\quad \times t(dx_k | x_{k-1}, u_{k-1}) \mu_{k-1}(du_{k-1} | x_0, u_0, \ldots, u_{k-2}, x_{k-1}) \cdots \\ &\quad \times t(dx_1 | x_0, u_0) \mu_0(du_0 | x_0) p(dx_0) \\ &= \int_{S_0 C_0 \ldots S_{k-1} C_{k-1}} \int_{\underline{S}_k} \mu_k(\underline{C}_k | x_0, u_0, \ldots, u_{k-1}, x_k) \\ &\quad \times t(dx_k | x_{k-1}, u_{k-1})\, dr_{k-1}(\pi, p) \qquad \forall \underline{S}_k \in \mathscr{B}_S, \underline{C}_k \in \mathscr{B}_C. \end{aligned} \tag{7}$$

From (1) and (7), we see that $q_k(\pi, p)(\Gamma) = 1$. If π is Markov, (7) becomes

$$q_k(\pi, p)(\underline{S}_k \underline{C}_k) = \int_{S_{k-1} C_{k-1}} \int_{\underline{S}_k} \mu_k(\underline{C}_k | x_k) t(dx_k | x_{k-1}, u_{k-1}) dq_{k-1}(\pi, p)$$
$$\forall \underline{S}_k \in \mathscr{B}_S, \quad \underline{C}_k \in \mathscr{B}_C. \tag{8}$$

If either

$$\int_{S_k C_k} g^- \, dq_k(\pi, p_x) < \infty \qquad \forall \pi \in \Pi', \quad x \in S, \quad k = 0, \ldots, N-1, \quad (\mathrm{F}^+)$$

or

$$\int_{S_k C_k} g^+ \, dq_k(\pi, p_x) < \infty \qquad \forall \pi \in \Pi', \quad x \in S, \quad k = 0, \ldots, N-1, \quad (\mathrm{F}^-)$$

then Lemma 7.11(b) implies that for every $\pi \in \Pi'$ and $x \in S$

$$J_{K,\pi}(x) = \sum_{k=0}^{K-1} \alpha^k \int_{S_k C_k} g \, dq_k(\pi, p_x), \qquad K = 1, \ldots, N. \tag{9}$$

If (F^+) [respectively (F^-)] appears preceding the statement of a proposition, then (F^+) [respectively (F^-)] is understood to be a part of the hypotheses of the proposition. If both (F^+) and (F^-) appear, then the proposition is valid when either (F^+) *or* (F^-) is included among the hypotheses.

If $\pi' \in \Pi'$ is a given policy, there may not exist a Markov policy which does at least as well as π' for every $x \in S$, i.e., a policy $\pi \in \Pi$ for which

$$J_{N,\pi}(x) \leq J_{N,\pi'}(x) \tag{10}$$

for every $x \in S$. However, if x is held fixed, then a Markov policy π can be found for which (10) holds.

Proposition 8.1 $(\mathrm{F}^+)(\mathrm{F}^-)$ If $x \in S$ and $\pi' \in \Pi'$, then there is a Markov policy π such that

$$J_{K,\pi}(x) = J_{K,\pi'}(x), \qquad K = 1, \ldots, N. \tag{11}$$

Proof Let $\pi' = (\mu'_0, \mu'_1, \ldots, \mu'_{N-1})$ be a policy and let $x \in S$ be given. For $k = 0, 1, \ldots, N-1$, let $\mu_k(du_k|x_k)$ be the Borel-measurable stochastic kernel obtained by decomposing $q_k(\pi', p_x)$ (Corollary 7.27.2), i.e.,

$$q_k(\pi', p_x)(\underline{S}_k \underline{C}_k) = \int_{\underline{S}_k} \mu_k(\underline{C}_k|x_k) p_k(\pi', p_x)(dx_k) \qquad \forall \underline{S}_k \in \mathscr{B}_S, \quad \underline{C}_k \in \mathscr{B}_C, \quad (12)$$

where $p_k(\pi', p_x)$ is the marginal of $q_k(\pi', p_k)$ on S_k. From (12) we see that

$$1 = q_k(\pi', p_x)(\Gamma) = \int_{S_k} \mu_k(U(x_k)|x_k) p_k(\pi', p_x)(dx_k),$$

so we must have $\mu_k(U(x_k)|x_k) = 1$ for $p_k(\pi', p_x)$ almost every x_k. By altering $\mu_k(\cdot|x_k)$ on a set of $p_k(\pi', p_x)$ measure zero if necessary, we may assume that (12) holds and $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ is a policy as set forth in Definition 8.2. In light of (9), (11) will follow if we show that $q_k(\pi', p_x) = q_k(\pi, p_x)$ for $k = 0, 1, \ldots, N-1$. For this, it suffices to show that, for $k = 0, 1, \ldots, N-1$,

$$q_k(\pi', p_x)(\underline{S}_k \underline{C}_k) = q_k(\pi, p_x)(\underline{S}_k \underline{C}_k) \qquad \forall \underline{S}_k \in \mathscr{B}_S, \quad \underline{C}_k \in \mathscr{B}_C. \quad (13)$$

We prove (13) by induction. For $k = 0$, $\underline{S}_0 \in \mathscr{B}_S$ and $\underline{C}_0 \in \mathscr{B}_C$, we have, from (12),

$$q_0(\pi', p_x)(\underline{S}_0 \underline{C}_0) = \int_{\underline{S}_0} \mu_0(\underline{C}_0|x_0) p_x(dx_0) = q_0(\pi, p_x)(\underline{S}_0 \underline{C}_0).$$

If $q_k(\pi', p_x) = q_k(\pi, p_x)$, then for $\underline{S}_{k+1} \in \mathscr{B}_S$, $\underline{C}_{k+1} \in \mathscr{B}_C$, we have, from (12),

$$q_{k+1}(\pi', p_x)(\underline{S}_{k+1} \underline{C}_{k+1}) = \int_{\underline{S}_{k+1}} \mu_{k+1}(\underline{C}_{k+1}|x_{k+1}) p_{k+1}(\pi', p_x)(dx_{k+1}). \quad (14)$$

From (7) we see that

$$p_{k+1}(\pi', p_x)(\underline{S}_{k+1}) = \int_{S_k C_k} t(\underline{S}_{k+1}|x_k, u_k)\, dq_k(\pi', p_x),$$

so if $h: S_{k+1} \to [0, \infty]$ is a Borel-measurable indicator function, then

$$\int_{S_{k+1}} h(x_{k+1})\, dp_{k+1}(\pi', p_x)(dx_{k+1}) = \int_{S_k C_k} \int_{S_{k+1}} h(x_{k+1}) t(dx_{k+1}|x_k, u_k) \times dq_k(\pi', p_x). \quad (15)$$

Then (15) holds for Borel-measurable simple functions, and finally, for all Borel-measurable functions $h: S_{k+1} \to [0, \infty]$. Letting $h(x_{k+1})$ in (15) be $\mu_{k+1}(\underline{C}_{k+1}|x_{k+1})$, we obtain from (14), the induction hypothesis, and (8)

$$\begin{aligned} q_{k+1}(\pi', p_x)(\underline{S}_{k+1} \underline{C}_{k+1}) &= \int_{S_k C_k} \int_{\underline{S}_{k+1}} \mu_{k+1}(\underline{C}_{k+1}|x_{k+1}) t(dx_{k+1}|x_k, u_k) \\ &\quad \times dq_k(\pi', p_x) \\ &= \int_{S_k C_k} \int_{\underline{S}_{k+1}} \mu_{k+1}(\underline{C}_{k+1}|x_{k+1}) t(dx_{k+1}|x_k, u_k)\, dq_k(\pi, p_x) \\ &= q_{k+1}(\pi, p_x)(\underline{S}_{k+1} \underline{C}_{k+1}), \end{aligned}$$

which proves (13) for $k+1$. Q.E.D.

Corollary 8.1.1 $(F^+)(F^-)$ For $K = 1, 2, \ldots, N$, we have

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

where Π is the set of all Markov policies.

Corollary 8.1.1 shows that the admission of non-Markov policies to our discussion has not resulted in a reduction of the optimal cost function. The advantage of allowing non-Markov policies is that an ε-optimal nonrandomized policy can then be guaranteed to exist (Proposition 8.3), whereas one may not exist within the class of Markov policies (Example 2).

8.2 The Dynamic Programming Algorithm—Existence of Optimal and ε-Optimal Policies

Let $U(C|S)$ denote the set of universally measurable stochastic kernels μ on C given S which satisfy $\mu(U(x)|x) = 1$ for every $x \in S$. Thus the set of Markov policies is $\Pi = U(C|S)U(C|S)\cdots U(C|S)$, where there are N factors.

Definition 8.4 Let $J: S \to R^*$ be universally measurable and $\mu \in U(C|S)$. The *operator* T_μ mapping J into $T_\mu(J): S \to R^*$ is defined by

$$T_\mu(J)(x) = \int_C [g(x,u) + \alpha \int_S J(x')t(dx'|x,u)]\mu(du|x)$$

for every $x \in S$.

The operator T_μ can also be written in terms of the system function f and the disturbance kernel $p(dw|x,u)$ as [cf. (3)]

$$T_\mu(J)(x) = \int_C [g(x,u) + \alpha \int_W J[f(x,u,w)]p(dw|x,u)]\mu(du|x).$$

By Proposition 7.46, $T_\mu(J)$ is universally measurable. We show that under (F^+) or (F^-), the cost corresponding to a policy $\pi = (\mu_0, \ldots, \mu_{N-1})$ can be defined in terms of the composition of operators $T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}}$.

Lemma 8.1 $(F^+)(F^-)$ Let $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ be a Markov policy and let $J_0: S \to R^*$ be identically zero. Then for $K = 1, 2, \ldots, N$ we have

$$J_{K,\pi} = (T_{\mu_0} \cdots T_{\mu_{K-1}})(J_0), \tag{16}$$

where $T_{\mu_0} \cdots T_{\mu_{K-1}}$ denotes the composition of $T_{\mu_0}, \ldots, T_{\mu_{K-1}}$.

Proof We proceed by induction. For $x \in S$,

$$\begin{aligned} J_{1,\pi}(x) &= \int g \, dq_0(\pi, p_x) \\ &= \int_{C_0} g(x, u_0)\mu_0(du_0|x) = T_{\mu_0}(J_0)(x). \end{aligned}$$

Suppose the lemma holds for $K-1$. Let $\bar{\pi}=(\mu_1,\mu_2,\ldots,\mu_{N-1},\mu)$, where μ is some element of $U(C|S)$. Then for any $x\in S$, the (F^+) or (F^-) assumption along with Lemma 7.11(b) implies that (5) can be rewritten as

$$\begin{aligned}J_{K,\pi}(x)&=\int_{C_0} g(x,u_0)\mu_0(du_0|x)+\alpha\int_{C_0}\int_{S_1}\int_{C_1}\cdots\int_{C_{K-1}}\left[\sum_{k=1}^{K-1}\alpha^{k-1}g(x_k,u_k)\right]\\&\quad\times\mu_{K-1}(du_{K-1}|x_{K-1})t(dx_{K-1}|x_{K-2},u_{K-2})\cdots\\&\quad\times\mu_1(du_1|x_1)t(dx_1|x,u_0)\mu_0(du_0|x)\\&=\int_{C_0} g(x,u_0)\mu_0(du_0|x)+\alpha\int_{C_0}\int_{S_1} J_{K-1,\bar{\pi}}(x_1)t(dx_1|x,u_0)\mu_0(du_0|x).\end{aligned}\tag{17}$$

Under (F^-), $\int_{C_0} g^+(x,u_0)\mu_0(du_0|x)<\infty$ and

$$\int_{C_0}\int_{S_1}[J_{K-1,\bar{\pi}}(x_1)t(dx_1|x,u_0)]^+\mu_0(du_0|x)<\infty,$$

while under (F^+) a similar condition holds, so Lemma 7.11(b) and the induction hypothesis can be applied to the right-hand side of (17) to conclude

$$\begin{aligned}J_{K,\pi}(x)&=\int_{C_0}\left[g(x,u_0)+\alpha\int_{S_1}(T_{\mu_1}\cdots T_{\mu_{K-1}})(J_0)(x_1)t(dx_1|x,u_0)\right]\mu_0(du_0|x)\\&=(T_{\mu_0}T_{\mu_1}\cdots T_{\mu_{K-1}})(J_0)(x).\qquad\text{Q.E.D.}\end{aligned}$$

Definition 8.5 Let $J:S\to R^*$ be universally measurable. The *operator* T mapping J into $T(J):S\to R^*$ is defined by

$$T(J)(x)=\inf_{u\in U(x)}\left\{g(x,u)+\alpha\int_S J(x')t(dx'|x,u)\right\}$$

for every $x\in S$.

Similarly as for T_μ, the operator T may be written in terms of f and $p(dw|x,u)$ as

$$T(J)(x)=\inf_{u\in U(x)}\left\{g(x,u)+\alpha\int_W J[f(x,u,w)]p(dw|x,u)\right\}.$$

If μ is nonrandomized, the operators T_μ and T of Definitions 8.4 and 8.5 are, except for measurability restrictions on J and μ, special cases of those defined in Section 2.1. In the present case, the mapping H of Section 2.1 is

$$\begin{aligned}H(x,u,J)&=g(x,u)+\alpha\int_S J(x')t(dx'|x,u)\\&=g(x,u)+\alpha\int_W J[f(x,u,w)]p(dw|x,u).\end{aligned}$$

We will state and prove versions of Assumptions F.1 and F.3 of Section 3.1 for this function H. Assumption F.2 is clearly true. Furthermore, if $\mu\in U(C|S)$,

$J_1, J_2 : S \to R^*$ are universally measurable, and $J_1 \leq J_2$, then $T_\mu(J_1) \leq T_\mu(J_2)$ and $T(J_1) \leq T(J_2)$. If $r \in (0, \infty)$, then $T_\mu(J_1 + r) = T_\mu(J_1) + \alpha r$ and $T(J_1 + r) = T(J_1) + \alpha r$. We will make frequent use of these properties. The reader should not be led to believe, however, that the model of this chapter is a special case of the model of Chapters 2 and 3. The earlier model does not admit measurability restrictions on policies.

By Lemma 7.30(4) and Propositions 7.47 and 7.48, $T(J)$ is lower semianalytic whenever J is. The composition of T with itself k times is denoted by T^k, i.e., $T^k(J) = T[T^{k-1}(J)]$, where $T^0(J) = J$. We show in Proposition 8.2 that under (F^+) or (F^-) the optimal cost can be defined in terms of T^N. Three preparatory lemmas are required.

Lemma 8.2 Let $J : S \to R^*$ be lower semianalytic. Then for $\varepsilon > 0$, there exists $\mu \in U(C|S)$ such that

$$T_\mu(J)(x) \leq T(J)(x) + \varepsilon \qquad \forall x \in S,$$

where $T(J)(x) + \varepsilon$ may be $-\infty$.

Proof By Proposition 7.50, there are universally measurable selectors $\mu_m : S \to C$ such that for $m = 1, 2, \ldots$ and $x \in S$, we have $\mu_m(x) \in U(x)$ and

$$T_{\mu_m}(J)(x) \leq \begin{cases} T(J)(x) + \varepsilon & \text{if } T(J)(x) > -\infty, \\ -2^m & \text{if } T(J)(x) = -\infty. \end{cases}$$

Let $\mu(du|x)$ assign mass one to $\mu_1(x)$ if $T(J)(x) > -\infty$ and assign mass $1/2^m$ to $\mu_m(x)$, $m = 1, 2, \ldots$, if $T(J)(x) = -\infty$.

For each $\underline{C} \in \mathscr{B}_C$,

$$\mu(\underline{C}|x) = \begin{cases} \chi_{\underline{C}}[\mu_1(x)] & \text{if } T(J)(x) > -\infty, \\ \sum_{m=1}^{\infty} (1/2^m)\chi_{\underline{C}}[\mu_m(x)] & \text{if } T(J)(x) = -\infty, \end{cases}$$

is a universally measurable function of x, and therefore μ is a universally measurable stochastic kernel [Lemma 7.28(a),(b)]. This μ has the desired properties. Q.E.D.

Lemma 8.3 (F^+) If $J_0 : S \to R^*$ is identically zero, then $T^K(J_0)(x) > -\infty$ for every $x \in S$, $K = 1, \ldots, N$, where T^K denotes the composition of T with itself K times.

Proof Suppose for some $K \leq N$ and $\bar{x} \in S$ that

$$T^j(J_0)(x) > -\infty, \qquad j = 0, \ldots, K-1,$$

for every $x \in S$, and

$$T^K(J_0)(\bar{x}) = -\infty.$$

By Proposition 7.50, there are universally measurable selectors $\mu_j: S \to C$, $j = 1, \ldots, K-1$, such that $\mu_j(x) \in U(x)$ and

$$(T_{\mu_{K-j}} T^{j-1})(J_0)(x) \leq T^j(J_0)(x) + 1, \qquad j = 1, \ldots, K-1,$$

for every $x \in S$. Then

$$\begin{aligned}(T_{\mu_1} \cdots T_{\mu_{K-1}})(J_0) &\leq (T_{\mu_1} \cdots T_{\mu_{K-2}})[T(J_0) + 1] \\ &\leq (T_{\mu_1} \cdots T_{\mu_{K-3}})[T^2(J_0) + 1 + \alpha] \\ &\leq T^{K-1}(J_0) + 1 + \alpha + \cdots + \alpha^{K-2},\end{aligned}$$

where the last inequality is obtained by repeating the process used to obtain the first two inequalities. By Lemma 8.2, there is a stochastic kernel $\mu_0 \in U(C|S)$ such that

$$(T_{\mu_0} T^{K-1})(J_0)(\bar{x}) = -\infty.$$

Then

$$(T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{K-1}})(J_0)(\bar{x}) \leq T_{\mu_0}[T^{K-1}(J_0) + 1 + \alpha + \cdots + \alpha^{K-2}](\bar{x}) = -\infty.$$

Choose any $\mu \in U(C|S)$ and let $\pi = (\mu_0, \ldots, \mu_{K-1}, \mu, \ldots, \mu)$, so that $\pi \in \Pi$. By Lemma 8.1,

$$\sum_{k=0}^{K-1} \alpha^k \int g \, dq_k(\pi, p_{\bar{x}}) = J_{K,\pi}(\bar{x}) = (T_{\mu_0} \cdots T_{\mu_{K-1}})(J_0)(\bar{x}) = -\infty,$$

so for some $k \leq K-1$, $\int g^- \, dq_k(\pi, p_{\bar{x}}) = \infty$. This contradicts the (F^+) assumption. Q.E.D.

Lemma 8.4 Let $\{J_k\}$ be a sequence of extended real-valued, universally measurable functions on S and let μ be an element of $U(C|S)$.

(a) If $T_\mu(J_1)(x) < \infty$ for every $x \in S$ and $J_k \downarrow J$, then $T_\mu(J_k) \downarrow T_\mu(J)$.

(b) If $T_\mu(J_1^-)(x) < \infty$ for every $x \in S$, $g \geq 0$, and $J_k \uparrow J$, then $T_\mu(J_k) \uparrow T_\mu(J)$.

(c) If $\{J_k\}$ is uniformly bounded, g is bounded, and $J_k \to J$, then $T_\mu(J_k) \to T_\mu(J)$.

Proof Assume first that $T_\mu(J_1) < \infty$ and $J_k \downarrow J$. Fix x. Since

$$\int [g(x,u) + \alpha \int J_1(x') t(dx'|x,u)] \mu(du|x) < \infty,$$

we have

$$g(x,u) + \alpha \int J_1(x') t(dx'|x,u) < \infty$$

for $\mu(du|x)$ almost all u. By the monotone convergence theorem [Lemma 7.11(f)],

$$g(x,u) + \alpha \int J_k(x') t(dx'|x,u) \downarrow g(x,u) + \alpha \int J(x') t(dx'|x,u)$$

for $\mu(du|x)$ almost all u. Apply the monotone convergence theorem again to conclude $T_\mu(J_k)(x) \downarrow T_\mu(J)(x)$.

If $T_\mu(J_1^-) < \infty$, $g \geq 0$, and $J_k \uparrow J$, the same type of argument applies. If $\{J_k\}$ is uniformly bounded, g bounded, and $J_k \to J$, a similar argument using the bounded convergence theorem applies. Q.E.D.

The dynamic programming algorithm over a finite horizon is executed by beginning with the identically zero function on S and applying the operator T successively N times. The next theorem says that this procedure generates the optimal cost function. In Proposition 8.3, we show how ε-optimal policies can also be obtained from this algorithm.

Proposition 8.2 $(F^+)(F^-)$ Let J_0 be the identically zero function on S. Then

$$J_K^* = T^K(J_0), \qquad K = 1, \ldots, N. \tag{18}$$

Proof It suffices to prove (18) for $K = N$, since the horizon N can be chosen to be any positive integer. For any $\pi = (\mu_0, \ldots, \mu_{N-1}) \in \Pi$ and $K \leq N$, we have

$$J_{K,\pi} = (T_{\mu_0} \cdots T_{\mu_{K-1}})(J_0) \geq (T_{\mu_0} \cdots T_{\mu_{K-2}} T)(J_0) \geq T^K(J_0), \tag{19}$$

where the last inequality is obtained by repeating the process used to obtain the first inequality. Infimizing over $\pi \in \Pi$ when $K = N$ and using Corollary 8.1.1, we obtain

$$J_N^* \geq T^N(J_0). \tag{20}$$

If (F^+) holds, then, by Lemma 8.3, $T^k(J_0) > -\infty$, $k = 1, \ldots, N$. For $\varepsilon > 0$, there are universally measurable selectors $\hat{\mu}_k : S \to C$, $k = 0, \ldots, N-1$, with $\hat{\mu}_k(x) \in U(x)$ and

$$\begin{aligned} -\infty < T_{\hat{\mu}_{N-k}}[T^{k-1}(J_0)](x) \\ \leq T^k(J_0)(x) + \varepsilon/(1 + \alpha + \alpha^2 + \cdots + \alpha^{N-1}), \qquad k = 1, \ldots, N, \end{aligned}$$

for every $x \in S$ (Proposition 7.50). Then

$$\begin{aligned} (T_{\hat{\mu}_0} T_{\hat{\mu}_1} \cdots T_{\hat{\mu}_{N-1}})(J_0) &\leq (T_{\hat{\mu}_0} T_{\hat{\mu}_1} \cdots T_{\hat{\mu}_{N-2}})[T(J_0) \\ &\quad + \varepsilon/(1 + \alpha + \alpha^2 + \cdots + \alpha^{N-1})] \\ &\leq (T_{\hat{\mu}_0} T_{\hat{\mu}_1} \cdots T_{\hat{\mu}_{N-3}})[T^2(J_0) \\ &\quad + \varepsilon(1 + \alpha)/(1 + \alpha + \alpha^2 + \cdots + \alpha^{N-1})] \\ &\leq T^N(J_0) + \varepsilon, \end{aligned} \tag{21}$$

where the last inequality is obtained by repeating the process used to obtain the first two inequalities. It follows that

$$J_N^* \leq T^N(J_0). \tag{22}$$

Combining (20) and (22), we see that the proposition holds under the (F^+) assumption.

If (F^-) holds, then $J_{K,\pi}(x) < \infty$ for every $x \in S$, $\pi \in \Pi$, $K = 1, \ldots, N$. Use Proposition 7.50 to choose nonrandomized policies $\pi^i = (\mu_0^i, \ldots, \mu_{N-1}^i) \in \Pi$ such that

$$(T_{\mu_k^i} T^{N-k-1})(J_0) \downarrow T^{N-k}(J_0), \qquad k = 0, \ldots, N-1,$$

as $i \to \infty$. By (19) and Lemma 8.4(a),

$$\begin{aligned} J_N^* &\leq \inf_{(i_0, \ldots, i_{N-1})} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-1}^{i_{N-1}}})(J_0) \\ &= \inf_{i_0} \cdots \inf_{i_{N-1}} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-1}^{i_{N-1}}})(J_0) \\ &= \inf_{i_0} \cdots \inf_{i_{N-2}} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-2}^{i_{N-2}}}) \left[\inf_{i_{N-1}} T_{\mu_{N-1}^{i_{N-1}}}(J_0) \right] \\ &= \inf_{i_0} \cdots \inf_{i_{N-2}} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-2}^{i_{N-2}}} T)(J_0) \\ &= T^N(J_0), \end{aligned} \tag{23}$$

where the last equality is obtained by repeating the process used to obtain the previous equality. Combining (20) and (23), we see that the proposition holds under the (F^-) assumption. Q.E.D.

When the state, control, and disturbance spaces are countable, the model of Definition 8.1 falls within the framework of Part I. Consider such a model, and, as in Part I, let M be the set of mappings $\mu: S \to C$ for which $\mu(x) \in U(x)$ for every $x \in S$. In Section 3.2, it was often assumed that for every $x \in S$ and $\mu_j \in M$, $j = 0, \ldots, K-1$, we have

$$(T_{\mu_0} \cdots T_{\mu_{K-1}})(J_0)(x) < \infty, \qquad K = 1, \ldots, N, \tag{24}$$

or else for every $x \in S$

$$\inf_{\mu_j \in M,\, 0 \leq j \leq K-1} (T_{\mu_0} \cdots T_{\mu_{K-1}})(J_0)(x) > -\infty, \qquad K = 1, \ldots, N. \tag{25}$$

Under the (F^+) assumption, Lemma 8.3 implies that

$$-\infty < T^K(J_0) \leq \inf_{\mu_j \in M,\, 0 \leq j \leq K-1} (T_{\mu_0} \cdots T_{\mu_{K-1}})(J_0),$$

so (25) is satisfied. Under (F^-), we have from Lemma 8.1 that

$$(T_{\mu_0} \cdots T_{\mu_{K-1}})(J_0) = J_{K,\pi} < \infty,$$

where $\pi = (\mu_0, \ldots, \mu_{K-1})$, so (24) holds. The primary reason for introducing the stronger (F^+) and (F^-) assumptions is to enable us to prove Lemma 8.1. If one chooses instead to take (16) as the definition of $J_{K,\pi}$ (as is done in

Section 3.1), then (24) or (25) suffices to prove Proposition 8.2 along the lines of the proof of Proposition 3.1 of Part I.

Proposition 8.2 implies the following property of the optimal cost function.

Corollary 8.2.1 $(F^+)(F^-)$ For $K = 1, 2, \ldots, N$, the function J_K^* is lower semianalytic.

Proof As observed following Definition 8.5, $T(J)$ is lower semianalytic whenever J is. Since $J_K^* = T^K(J_0)$ and $J_0 \equiv 0$ is lower semianalytic, the result follows. Q.E.D.

We give an example to show that even when $\Gamma = SC$ and the one-stage cost $g: SC \to R^*$ is Borel-measurable, J_1^* can fail to be Borel-measurable.

EXAMPLE 1 Let A be an analytic subset of $[0, 1]$ which is not Borel-measurable (Appendix B). By Proposition 7.39, there is a closed set $F \subset [0,1]\mathcal{N}$ such that $A = \text{proj}_{[0,1]}(F)$. Let $S = [0,1]$, $C = \mathcal{N}$, $\Gamma = SC$, and $g = \chi_{F^c}$. Then

$$J_1^*(x) = \inf_{u \in C} g(x,u) = \chi_{A^c}(x) \qquad \forall x \in S,$$

which is a lower semianalytic but not Borel-measurable function. We could also choose $C = [0,1]$, $\Gamma = SC$, B a G_δ-subset of the unit square SC, and $g = \chi_{B^c}$. This is because $\mathcal{N}$ and $\mathcal{N}_0$, the space of irrational numbers in $[0,1]$, are homeomorphic (Proposition 7.5). But

$$\mathcal{N}_0 = \bigcap_{r \in Q} ([0,1] - \{r\})$$

is a G_δ-subset of $[0,1]$, so there is a homeomorphism $\varphi: \mathcal{N} \to [0,1]$ such that $\varphi(\mathcal{N})$ is a G_δ-subset of $[0,1]$. Let $\Phi: [0,1]\mathcal{N} \to [0,1][0,1]$ be the homeomorphism defined by

$$\Phi(x,z) = (x, \varphi(z)).$$

Then $\Phi([0,1]\mathcal{N}) = [0,1]\varphi(\mathcal{N})$ is a G_δ-subset of $SC = [0,1][0,1]$, and since F is a G_δ-set in $[0,1]\mathcal{N}$, $B = \Phi(F)$ is a G_δ-subset of SC which satisfies $\text{proj}_S(B) = A$. If $g = \chi_{B^c}$, then again $J_1^* = \chi_{A^c}$.

We now use Proposition 8.2 to establish existence of ε-optimal policies.

Proposition 8.3

(F^+) For each $\varepsilon > 0$, there exists a nonrandomized Markov ε-optimal policy.

(F^-) For each $\varepsilon > 0$, there exists a nonrandomized semi-Markov ε-optimal policy and a (randomized) Markov ε-optimal policy.

Proof If (F^+) holds, then the policy $(\hat{\mu}_0, \ldots, \hat{\mu}_{N-1})$ constructed in the proof of Proposition 8.2 is ε-optimal, nonrandomized, and Markov.

Assume (F^-) holds. We show first the existence of an ε-optimal, nonrandomized, semi-Markov policy. Let $\pi^i = (\mu_0^i, \ldots, \mu_{N-1}^i)$ be as in the proof of Proposition 8.2. Then

$$J_N^* = T^N(J_0) = \inf_{(i_0, \ldots, i_{N-1})} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-1}^{i_{N-1}}})(J_0)$$
$$= \inf_{(i_0, \ldots, i_{N-1})} J_{N, \pi^{(i_0, \ldots, i_{N-1})}},$$

where $\pi^{(i_0, \ldots, i_{N-1})} = (\mu_0^{i_0}, \ldots, \mu_{N-1}^{i_{N-1}})$. Choose $\varepsilon > 0$ and define

$$\theta(x) = \begin{cases} J_N^*(x) + \varepsilon & \text{if } J_N^*(x) > -\infty, \\ -1/\varepsilon & \text{if } J_N^*(x) = -\infty. \end{cases}$$

Order linearly the countable set $\{\pi^{(i_0, \ldots, i_{N-1})} | i_0, \ldots, i_{N-1}$ are positive integers$\}$ and define $\pi(x)$ to be the first $\pi^{(i_0, \ldots, i_{N-1})}$ such that

$$J_{N, \pi^{(i_0, \ldots, i_{N-1})}}(x) \leq \theta(x).$$

Let the components of $\pi(x)$ be

$$(\mu_0^x(x_0), \mu_1^x(x_1), \ldots, \mu_{N-1}^x(x_{N-1})).$$

The set $\{x | \pi(x) = \pi^{(i_0, \ldots, i_{N-1})}\}$ is universally measurable for each $(i_0, \ldots, i_{N-1})$, so

$$(\mu_0(x_0), \mu_1(x_0, x_1), \ldots, \mu_{N-1}(x_0, x_{N-1})),$$

where $\mu_0(x_0) = \mu_0^{x_0}(x_0)$ and $\mu_k(x_0, x_k) = \mu_k^{x_0}(x_k)$, $k = 1, \ldots, N-1$, is an ε-optimal nonrandomized semi-Markov policy.

We now show the existence of an ε-optimal (randomized) Markov policy. By Lemma 8.2, there exist $\mu_{N-k} \in U(C|S)$ such that for $k = 1, \ldots, N$

$$(T_{\mu_{N-k}} T^{k-1})(J_0) \leq T^k(J_0) + \varepsilon/(1 + \alpha + \alpha^2 + \cdots + \alpha^{N-1}).$$

Proceed as in (21). Q.E.D.

If the (F^-) assumption holds and $\varepsilon > 0$, it may not be possible to find a nonrandomized Markov ε-optimal policy, as the following example demonstrates.

EXAMPLE 2 Let $S = \{0, 1, 2, \ldots\}$, $C = \{1, 2, \ldots\}$, $W = \{w_1, w_2\}$, $\Gamma = SC$, $N = 2$ and define

$$g(x, u) = \begin{cases} -u & \text{if } x = 1, \\ 0 & \text{if } x \neq 1, \end{cases}$$

$$f(x, u, w) = \begin{cases} 0 & \text{if } x = 0 \text{ or } x = 1 \text{ or } w = w_1, \\ 1 & \text{if } x \neq 0,\ x \neq 1, \text{ and } w = w_2, \end{cases}$$

$$p(\{w_1\} | x, u) = 1 - 1/x \quad \text{if } x \neq 0, \quad x \neq 1,$$
$$p(\{w_2\} | x, u) = 1/x \quad \text{if } x \neq 0, \quad x \neq 1.$$

The (F^-) assumption is satisfied. Let $\pi = (\mu_0, \mu_1)$ be a nonrandomized Markov policy. If the initial state x_0 is neither zero nor one, then regardless of the policy employed, $x_1 = 0$ with probability $1 - (1/x_0)$, and $x_1 = 1$ with probability $1/x_0$. Once the system reaches zero, it remains there at no further cost. If the system reaches one, it moves to $x_2 = 0$ at a cost of $-\mu_1(1)$. Thus $J_{N,\pi}(x_0) = -\mu_1(1)/x_0$ if $x_0 \neq 0$, $x_0 \neq 1$, and $J_N^*(x_0) = -\infty$ if $x_0 \neq 0$, $x_0 \neq 1$. For any $\varepsilon > 0$, π cannot be ε-optimal.

In Example 2, it is possible to find a sequence of nonrandomized Markov policies $\{\pi_n\}$ such that $J_{N,\pi_n} \downarrow J_N^*$. This example motivates the idea of policies exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality (Definition 8.3) and Proposition 8.4, which we prove with the aid of the next lemma.

Lemma 8.5 Let $\{J_k\}$ be a sequence of universally measurable functions from S to R^* and μ a universally measurable function from S to C whose graph lies in Γ. Suppose for some sequence $\{\varepsilon_k\}$ of positive numbers with $\sum_{k=1}^{\infty} \varepsilon_k < \infty$, we have, for every $x \in S$,

$$\int J_1^+(x')t(dx'|x, \mu(x)) < \infty, \qquad \lim_{k\to\infty} J_k(x) = J(x)$$

and for $k = 2, 3, \ldots$

$$\begin{aligned} J(x) \leq J_k(x) &\leq J(x) + \varepsilon_k \qquad \text{if} \quad J(x) > -\infty, \\ J_k(x) &\leq J_{k-1}(x) + \varepsilon_k \qquad \text{if} \quad J(x) = -\infty. \end{aligned}$$

Then

$$\lim_{k\to\infty} T_\mu(J_k) = T_\mu(J). \tag{26}$$

Proof Since $J \leq J_k$ for every k, it is clear that

$$T_\mu(J) \leq \liminf_{k\to\infty} T_\mu(J_k). \tag{27}$$

For $x \in S$,

$$\limsup_{k\to\infty} T_\mu(J_k)(x) \leq g[x, \mu(x)] + \alpha \limsup_{k\to\infty} \int_{\{x'|J(x')>-\infty\}} J_k(x')t(dx'|x, \mu(x))$$
$$+ \alpha \limsup_{k\to\infty} \int_{\{x'|J(x')=-\infty\}} J_k(x')t(dx'|x, \mu(x)).$$

Now

$$\limsup_{k\to\infty} \int_{\{x'|J(x')>-\infty\}} J_k(x')t(dx'|x, \mu(x))$$
$$\leq \limsup_{k\to\infty} \left[\int_{\{x'|J(x')>-\infty\}} J(x')t(dx'|x, \mu(x)) + \varepsilon_k \right]$$
$$= \int_{\{x'|J(x')>-\infty\}} J(x')t(dx'|x, \mu(x)).$$

If $J(x') = -\infty$, then

$$J_k(x') + \sum_{n=k+1}^{\infty} \varepsilon_n \downarrow J(x'),$$

and since $\int [J_1^+(x') + \sum_{n=2}^{\infty} \varepsilon_n] t(dx'|x, \mu(x)) < \infty$, we have

$$\begin{aligned}
&\limsup_{k\to\infty} \int_{\{x'|J(x')=-\infty\}} J_k(x') t(dx'|x, \mu(x)) \\
&\quad \le \lim_{k\to\infty} \int_{\{x'|J(x')=-\infty\}} \left[J_k(x') + \sum_{n=k+1}^{\infty} \varepsilon_n \right] t(dx'|x, \mu(x)) \\
&\quad = \int_{\{x'|J(x')=-\infty\}} J(x') t(dx'|x, \mu(x)).
\end{aligned}$$

If follows that

$$\limsup_{k\to\infty} T_\mu(J_k) \le T_\mu(J). \tag{28}$$

Combine (27) and (28) to conclude (26). Q.E.D.

Proposition 8.4 (F^-) Let $\{\varepsilon_n\}$ be a sequence of positive numbers with $\varepsilon_n \downarrow 0$. There exists a sequence of nonrandomized Markov policies $\{\pi_n\}$ exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality. In particular, if $J_N^*(x) > -\infty$ for all $x \in S$, then for every $\varepsilon > 0$ there exists an ε-optimal nonrandomized Markov policy.

Proof For $N = 1$, by Proposition 7.50 there exists a sequence of nonrandomized Markov policies $\pi_n = (\mu_0^n)$ such that for all n

$$T_{\mu_0^n}(J_0)(x) \le \begin{cases} T(J_0)(x) + \varepsilon_n & \text{if } T(J_0)(x) > -\infty, \\ -1/\varepsilon_n & \text{if } T(J_0)(x) = -\infty. \end{cases}$$

We may assume without loss of generality that

$$T_{\mu_0^n}(J_0) \le T_{\mu_0^{n-1}}(J_0).$$

Therefore $\{\pi_n\}$ exhibits $\{\varepsilon_n\}$-dominated convergence to one-stage optimality.

Suppose the result holds for $N - 1$. Let $\pi_n = (\mu_1^n, \ldots, \mu_{N-1}^n)$ be a sequence of $(N-1)$-stage nonrandomized Markov policies exhibiting $\{\varepsilon_n/2\alpha\}$-dominated convergence to $(N-1)$-stage optimality, i.e.,

$$\lim_{n\to\infty} J_{N-1,\pi_n} = J_{N-1}^*,$$

$$J_{N-1,\pi_n}(x) \le \begin{cases} J_{N-1}^*(x) + (\varepsilon_n/2\alpha) & \text{if } J_{N-1}^*(x) > -\infty, \quad (29) \\ J_{N-1,\pi_{n-1}}(x) + (\varepsilon_n/2\alpha) & \text{if } J_{N-1}^*(x) = -\infty. \quad (30) \end{cases}$$

We assume without loss of generality that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. By Proposition 7.50, there exists a sequence $\{\mu^n\}$ of universally measurable functions from S to

C whose graphs lie in Γ such that

$$T_{\mu^n}(J^*_{N-1})(x) \leq \begin{cases} J^*_N(x) + (\varepsilon_n/2) & \text{if} \quad J^*_N(x) > -\infty, \\ -2/\varepsilon_n & \text{if} \quad J^*_N(x) = -\infty. \end{cases} \tag{31}$$

We may assume without loss of generality that

$$T_{\mu^n}(J^*_{N-1}) \leq T_{\mu^{n-1}}(J^*_{N-1}), \qquad n = 2, 3, \ldots. \tag{32}$$

By Proposition 7.48, the set

$$\begin{aligned} A(J^*_{N-1}) &= \{(x,u) \in \Gamma \,|\, t(\{x' \,|\, J^*_{N-1}(x') = -\infty\} \,|\, x, u) > 0\} \\ &= \left\{(x,u) \in \Gamma \,\middle|\, \int -\chi_{\{x' \,|\, J^*_{N-1}(x') = -\infty\}}(x') t(dx' \,|\, x, u) < 0 \right\} \end{aligned} \tag{33}$$

is analytic in SC, and the Jankov-von Neumann theorem (Proposition 7.49) implies the existence of a universally measurable $\mu : \text{proj}_S[A(J^*_{N-1})] \to C$ whose graph lies in $A(J^*_{N-1})$. Define

$$\hat{\mu}^n(x) = \begin{cases} \mu(x) & \text{if} \quad x \in \text{proj}_S[A(J^*_{N-1})], \\ \mu^n(x) & \text{otherwise.} \end{cases}$$

Then $\hat{\pi}_n = (\hat{\mu}^n, \pi_n)$ is an N-stage nonrandomized Markov policy which will be shown to exhibit $\{\varepsilon_n\}$-dominated convergence to optimality.

For $x \in \text{proj}_S[A(J^*_{N-1})]$, we have, from Lemma 8.5 and the choice of μ,

$$\begin{aligned} \limsup_{n \to \infty} J_{N, \hat{\pi}_n}(x) &= \limsup_{n \to \infty} T_{\mu}(J_{N-1, \pi_n})(x) \\ &= T_{\mu}(J^*_{N-1})(x) = -\infty. \end{aligned}$$

For $x \notin \text{proj}_S[A(J^*_{N-1})]$, we have $t(\{x' \,|\, J^*_{N-1}(x') = -\infty\} \,|\, x, u) = 0$ for every $u \in U(x)$, so by (29)

$$J_{N, \hat{\pi}_n}(x) = T_{\mu^n}(J_{N-1, \pi_n})(x) \leq T_{\mu^n}(J^*_{N-1})(x) + \varepsilon_n/2, \tag{34}$$

and

$$\limsup_{n \to \infty} J_{N, \hat{\pi}_n}(x) \leq \limsup_{n \to \infty} T_{\mu^n}(J^*_{N-1})(x) \leq J^*_N(x)$$

by (31). It follows that

$$\lim_{n \to \infty} J_{N, \hat{\pi}_n} = J^*_N. \tag{35}$$

Suppose for fixed $x \in S$ we have $J^*_N(x) > -\infty$. Then $x \notin \text{proj}_S[A(J^*_{N-1})]$, and we have, from (31) and (34),

$$J_{N, \hat{\pi}_n}(x) \leq T_{\mu^n}(J^*_{N-1})(x) + \varepsilon_n/2 \leq J^*_N(x) + \varepsilon_n. \tag{36}$$

Suppose now that $J_N^*(x) = -\infty$. If $x \notin \text{proj}_S[A(J_{N-1}^*)]$, then (32) and (34) imply, for $n \geq 2$,

$$\begin{aligned} J_{N,\hat{\pi}_n}(x) &\leq T_{\mu^n}(J_{N-1}^*)(x) + \varepsilon_n/2 \\ &\leq T_{\mu^{n-1}}(J_{N-1}^*)(x) + \varepsilon_n/2 \\ &\leq T_{\mu^{n-1}}(J_{N-1,\pi_{n-1}})(x) + \varepsilon_n/2 \\ &\leq J_{N,\hat{\pi}_{n-1}}(x) + \varepsilon_n/2, \end{aligned}$$

while if $x \in \text{proj}_S[A(J_{N-1}^*)]$, we have, from (29) and (30),

$$\begin{aligned} J_{N,\hat{\pi}_n}(x) &= T_\mu(J_{N-1,\pi_n})(x) \\ &\leq T_\mu(J_{N-1,\pi_{n-1}})(x) + \varepsilon_n/2 \\ &= J_{N,\hat{\pi}_{n-1}}(x) + \varepsilon_n/2. \end{aligned}$$

In either case,

$$J_{N,\hat{\pi}_n}(x) \leq J_{N,\hat{\pi}_{n-1}}(x) + \varepsilon_n. \tag{37}$$

From (35)–(37) we see that $\{\hat{\pi}_n\}$ exhibits $\{\varepsilon_n\}$-dominated convergence to optimality. Q.E.D.

We conclude our discussion of the ramifications of Proposition 8.2 with a technical result needed for the development in Chapter 10.

Lemma 8.6 $(F^+)(F^-)$ For every $p \in P(S)$,

$$\int J_N^*(x)p(dx) = \inf_{\pi \in \Pi} \int J_{N,\pi}(x)p(dx).$$

Proof For $p \in P(S)$ and $\pi \in \Pi$,

$$\int J_N^*(x)p(dx) \leq \int J_{N,\pi}(x)p(dx),$$

which implies

$$\int J_N^*(x)p(dx) \leq \inf_{\pi \in \Pi} \int J_{N,\pi}(x)p(dx). \tag{38}$$

Choose $\varepsilon > 0$ and let $\hat{\pi} \in \Pi$ be ε-optimal. If $p(\{x | J_N^*(x) = -\infty\}) = 0$, then

$$\int J_{N,\hat{\pi}}(x)p(dx) \leq \int J_N^*(x)p(dx) + \varepsilon,$$

and it follows that

$$\inf_{\pi \in \Pi} \int J_{N,\pi}(x)p(dx) \leq \int J_N^*(x)p(dx). \tag{39}$$

If $p(\{x|J_N^*(x) = -\infty\}) > 0$, then

$$\int J_{N,\tilde{\pi}}(x)p(dx) \leq -p(\{x|J_N^*(x) = -\infty\})/\varepsilon + \int_{\{x|J_N^*(x) > -\infty\}} J_N^*(x)p(dx) + \varepsilon. \tag{40}$$

If $\int_{\{x|J_N^*(x)>-\infty\}} J_N^*(x)p(dx) = \infty$, then $\int J_N^*(x)p(dx) = \infty$ and (39) follows. Otherwise, the right-hand side of (40) can be made arbitrarily small by letting ε approach zero, so $\inf_{\pi\in\Pi}\int J_{N,\pi}(x)p(dx) = -\infty$ and (39) is again valid. The lemma follows from (38) and (39). Q.E.D.

We now consider the question of constructing an optimal policy, if this is at all possible. When the dynamic programming algorithm can be used to construct an optimal policy, this policy usually satisfies a condition stronger than mere optimality. This condition is given in the next definition.

Definition 8.6 Let $\pi = (\mu_0, \ldots, \mu_{N-1})$ be a Markov policy and $\pi^{N-k} = (\mu_k, \ldots, \mu_{N-1})$, $k = 0, \ldots, N-1$. The policy π is *uniformly N-stage optimal* if

$$J_{N-k,\pi^{N-k}} = J_{N-k}^*, \qquad k = 0, \ldots, N-1.$$

Lemma 8.7 $(F^+)(F^-)$ The policy $\pi = (\mu_0, \ldots, \mu_{N-1}) \in \Pi$ is uniformly N-stage optimal if and only if

$$(T_{\mu_k}T^{N-k-1})(J_0) = T^{N-k}(J_0), \qquad k = 0, \ldots, N-1.$$

Proof If $\pi = (\mu_0, \ldots, \mu_{N-1})$ is uniformly N-stage optimal, then

$$\begin{aligned} T^{N-k}(J_0) = J_{N-k}^* = J_{N-k,\pi^{N-k}} &= T_{\mu_k}(J_{N-k-1,\pi^{N-k-1}}) \\ &= T_{\mu_k}(J_{N-k-1}^*) = (T_{\mu_k}T^{N-k-1})(J_0), \qquad k = 0, \ldots, N-1, \end{aligned}$$

where $J_{0,\pi^0} = J_0^* \equiv 0$. If $(T_{\mu_k}T^{N-k-1})(J_0) = T^{N-k}(J_0)$, $k = 0, \ldots, N-1$, then for all k

$$\begin{aligned} J_{N-k}^* = T^{N-k}(J_0) &= (T_{\mu_k}T^{N-k-1})(J_0) \\ &= (T_{\mu_k}T_{\mu_{k+1}}T^{N-k-2})(J_0) \\ &= (T_{\mu_k}\cdots T_{\mu_{N-1}})(J_0) \\ &= J_{N-k,\pi^{N-k}}, \end{aligned}$$

where the next to last equality is obtained by continuing the process used to obtain the previous equalities. Q.E.D.

Lemma 8.7 is the analog for the Borel model of Proposition 3.3 for the model of Part I. Because (F^+) or (F^-) is a required assumption in Lemma 8.1, one of them is also required in Lemma 8.7, as the following example shows. If we take (16) as the definition of $J_{k,\pi}$, then Lemma 8.7, Proposition 8.5, and

Corollaries 8.5.1 and 8.5.2 hold without the (F^+) and (F^-) assumptions. The proofs are similar to those of Section 3.2.

EXAMPLE 3 Let $S = \{s,t\} \cup \{(k,j) | k = 1,2,\ldots;\ j = 1,2\}$, $C = \{a,b\}$, $U(s) = \{a,b\}$, $U(t) = U(k,j) = \{b\}$, $k = 1,2,\ldots, j = 1,2$, $W = S$, and $\alpha = 1$. Let the disturbance kernel be given by $p(s|s,a) = 1$,

$$p[(k,1)|s,b] = p[(k,2)|(l,1),b] = k^{-2}\left(\sum_{n=1}^{\infty} \frac{1}{n^2}\right)^{-1}, \qquad k,l = 1,2,\ldots,$$

$p[t|(k,2),b] = 1$, $k = 1,2,\ldots$, and $p(t|t,b) = 1$. Let the system function be $f(x,u,w) = w$. Thus if the system begins at state s, we can hold it at s or allow it to move to some state $(l,1)$, from which it subsequently moves to some $(k,2)$ and then to t. Having reached t, the system remains there. The relevant costs are $g(s,a) = g(s,b) = g(t,b) = 0$, $g[(k,1),b] = k$, $g[(k,2),b] = -k$, $k = 1,2,\ldots$. Let $\pi = (\mu_0,\mu_1,\mu_2)$ be a policy with $\mu_0(s) = b$, $\mu_1(s) = \mu_2(s) = a$. Then

$$T(J_0)(x_2) = T_{\mu_2}(J_0)(x_2) = \begin{cases} 0 & \text{if } x_2 = s, \\ k & \text{if } x_2 = (k,1), \\ -k & \text{if } x_2 = (k,2), \\ 0 & \text{if } x_2 = t, \end{cases}$$

$$T^2(J_0)(x_1) = (T_{\mu_1}T_{\mu_2})(J_0)(x_1) = \begin{cases} 0 & \text{if } x_1 = s, \\ -\infty & \text{if } x_1 = (k,1), \\ -k & \text{if } x_1 = (k,2), \\ 0 & \text{if } x_1 = t, \end{cases}$$

$$T^3(J_0)(x_0) = (T_{\mu_0}T_{\mu_1}T_{\mu_2})(J_0)(x_0) = \begin{cases} -\infty & \text{if } x_0 = s, \\ -\infty & \text{if } x_0 = (k,1), \\ -k & \text{if } x_0 = (k,2), \\ 0 & \text{if } x_0 = t. \end{cases}$$

However, $J_{\pi,3}(s) = \infty > J_{\bar{\pi},3}(s) = 0$, where $\bar{\pi} = (\bar{\mu}_0,\bar{\mu}_1,\bar{\mu}_2)$ and $\bar{\mu}_0(s) = \bar{\mu}_1(s) = \bar{\mu}_2(s) = a$, so π is not optimal and $T^3(J_0) \neq J_3^*$. It is easily verified that $\bar{\pi}$ is a uniformly three-stage optimal policy, so Corollary 3.3.1(b) also fails to hold for the Borel model of this chapter. Here both assumptions (F^+) and (F^-) are violated.

Proposition 8.5 $(F^+)(F^-)$ If the infimum in

$$\inf_{u \in U(x)} \left\{ g(x,u) + \alpha \int J_k^*(x') t(dx'|x,u) \right\}, \qquad k = 0,\ldots,N-1, \tag{41}$$

is achieved for each $x \in S$, where J_0^* is identically zero, then a uniformly N-stage optimal (and hence optimal) nonrandomized Markov policy exists. This policy is generated by the dynamic programming algorithm, i.e., by measurably selecting for each x a control u which achieves the infimum.

Proof Let $\pi = (\mu_0, \ldots, \mu_{N-1})$, where $\mu_{N-k-1}: S \to C$ achieves the infimum in (41) and satisfies $\mu_{N-k-1}(x) \in U(x)$ for every $x \in S, k = 0, \ldots, N-1$ (Proposition 7.50). Apply Lemma 8.7. Q.E.D.

Corollary 8.5.1 $(F^+)(F^-)$ If $U(x)$ is a finite set for each $x \in S$, then a uniformly N-stage optimal nonrandomized Markov policy exists.

Corollary 8.5.2 $(F^+)(F^-)$ If for each $x \in S$, $\lambda \in R$, and $k = 0, \ldots, N-1$, the set

$$U_k(x, \lambda) = \left\{ u \in U(x) \middle| g(x, u) + \alpha \int J_k^*(x') t(dx' | x, u) \leq \lambda \right\}$$

is compact, then there exists a uniformly N-stage optimal nonrandomized Markov policy.

Proof Apply Lemma 3.1 to Proposition 8.5. Q.E.D.

8.3 The Semicontinuous Models

Along the lines of our development of lower and upper semicontinuous functions in Section 7.5, we can consider lower and upper semicontinuous decision models. Our models will be designed to take advantage of the possibility for Borel-measurable selection (Propositions 7.33 and 7.34), and in the case of lower semicontinuity, the attainment of the infimum in (55) of Chapter 7. We discuss the lower semicontinuous model first.

Definition 8.7 The *lower semicontinuous, finite horizon, stochastic, optimal control model* is a nine-tuple $(S, C, U, W, p, f, \alpha, g, N)$ as given in Definition 8.1 which has the following additional properties:

(a) The control space C is compact.

(b) The set Γ defined by (1) has the form $\Gamma = \bigcup_{j=1}^{\infty} \Gamma^j$, where $\Gamma^1 \subset \Gamma^2 \subset \cdots$, each Γ^j is a closed subset of SC, and

$$\lim_{j \to \infty} \inf_{(x,u) \in \Gamma^j - \Gamma^{j-1}} g(x, u) = \infty.^{\dagger}$$

(c) The disturbance kernel $p(dw|x, u)$ is continuous on Γ.

† By convention, the infimum over the empty set is ∞, so this condition is satisfied if the Γ^j are all identical for j larger than some index k.

(d) The system function f is continuous on ΓW.

(e) The one-stage cost function g is lower semicontinuous and bounded below on Γ.

Conditions (c) and (d) of Definition 8.7 and Proposition 7.30 imply that $t(dx'|x,u)$ defined by (3) is continuous on Γ, since for any $h \in C(S)$ we have

$$\int h(x')t(dx'|x,u) = \int h[f(x,u,w)]p(dw|x,u).$$

Condition (e) implies that the (F^+) assumption holds.

Proposition 8.6 Consider the lower semicontinuous finite horizon model of Definition 8.7. For $k = 1,2,\ldots,N$, the k-stage optimal cost function J_k^* is lower semicontinuous and bounded below, and $J_k^* = T^k(J_0)$. Furthermore, a Borel-measurable, uniformly N-stage optimal, nonrandomized Markov policy exists.

Proof Suppose $J: S \to R^*$ is lower semicontinuous and bounded below, and define $K: \Gamma \to R^*$ by

$$K(x,u) = g(x,u) + \alpha \int J(x')t(dx'|x,u). \tag{42}$$

Extend K to all of SC by defining

$$\hat{K}(x,u) = \begin{cases} K(x,u) & \text{if } (x,u) \in \Gamma, \\ \infty & \text{if } (x,u) \notin \Gamma. \end{cases}$$

By Proposition 7.31(a) and the remarks following Lemma 7.13, the function K is lower semicontinuous on Γ. For $c \in R$, the set $\{(x,u) \in SC | \hat{K}(x,u) \leq c\}$ must be contained in some Γ^k by Definition 7.8(b), so the set

$$\{(x,u) \in SC | \hat{K}(x,u) \leq c\} = \{(x,u) \in \Gamma^k | K(x,u) \leq c\}$$

is closed in Γ^k and thus closed in SC as well. It follows that $\hat{K}(x,u)$ is lower semicontinous and bounded below on SC and, by Proposition 7.32, the function

$$T(J)(x) = \inf_{u \in C} \hat{K}(x,u) \tag{43}$$

is as well. In fact, Proposition 7.33 states that the infimum in (43) is achieved for every $x \in S$, and there exists a Borel-measurable $\varphi: S \to C$ such that

$$T(J)(x) = \hat{K}[x, \varphi(x)] \qquad \forall x \in S.$$

For $j = 1,2,\ldots,$ let φ_j: $\text{proj}_S(\Gamma^j) \to C$ be a Borel-measurable function with graph in Γ^j. (Set $D = \Gamma^j$ in Proposition 7.33 to establish the existence of such a function.) Define $\mu: S \to C$ so that $\mu(x) = \varphi(x)$ if $T(J)(x) < \infty$, $\mu(x) = \varphi_1(x)$ if $T(J)(x) = \infty$ and $x \in \text{proj}_S(\Gamma^1)$; and for $j = 2,3,\ldots,$ define $\mu(x) = \varphi_j(x)$ if

$T(J)(x) = \infty$ and $x \in \text{proj}_S(\Gamma^j) - \text{proj}_S(\Gamma^{j-1})$. Then μ is Borel-measurable, $\mu(x) \in U(x)$ for every $x \in S$, and $T_\mu(J) = T(J)$.

Since $J_0 \equiv 0$ is lower semicontinuous and bounded below, the above argument shows that $J_k^* = T^k(J_0)$ has these properties also, and furthermore, for each $k = 0, \ldots, N-1$, there exists a Borel-measurable $\mu_k : S \to C$ such that $\mu_k(x) \in U(x)$ for every $x \in S$ and $(T_{\mu_k} T^{N-k-1})(J_0) = T^{N-k}(J_0)$. The proposition follows from Lemma 8.7. Q.E.D.

We note that although condition (a) of Definition 8.7 requires the compactness of C, the conclusion of Proposition 8.6 still holds if C is not compact but can be homeomorphically embedded in a compact space $\hat{C}$ in such a way that the image of Γ^j, $j = 1, 2, \ldots$, is closed in $S\hat{C}$. That is to say, the conclusion holds if there is a compact space $\hat{C}$ and a homeomorphism $\varphi : C \to \hat{C}$ such that for $j = 1, 2, \ldots$, $\Phi(\Gamma^j)$ is closed in $S\hat{C}$, where

$$\Phi(x, u) = (x, \varphi(u)).$$

The continuity of f and $p(dw|x, u)$ and the lower semicontinuity of g are unaffected by this embedding. In particular, if Γ^j is compact for each j, we can take $\hat{C} = \mathscr{H}$ and use Urysohn's theorem (Proposition 7.2) and the fact that the continuous image of a compact set is compact to accomplish this transformation. We state this last result as a corollary.

Corollary 8.6.1 The conclusions of Proposition 8.6 hold if instead of assuming that C is compact and each Γ^j is closed in Definition 8.7, we assume that each Γ^j is compact.

Definition 8.8 The *upper semicontinuous, finite horizon, stochastic, optimal control model* is a nine-tuple $(S, C, U, W, p, f, \alpha, g, N)$ as given in Definition 8.1 which has the following additional properties:

(a) The set Γ defined by (1) is open in SC.
(b) The disturbance kernel $p(dw|x, u)$ is continuous on Γ.
(c) The system function f is continuous on ΓW
(d) The one-stage cost g is upper semicontinuous and bounded above on Γ.

As in the lower semicontinuous model, the stochastic kernel $t(dx'|x, u)$ is continuous in the upper semicontinuous model. In the upper semicontinuous model, the (F^-) assumption holds. If $J : S \to R^*$ is upper semicontinuous and bounded above, then $K : \Gamma \to R^*$ defined by (42) is upper semicontinuous and bounded above. By Proposition 7.34, the function

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

is upper semicontinuous, and for every $\varepsilon > 0$ there exists a Borel-measurable

$\mu: S \to C$ such that $\mu(x) \in U(x)$ for every $x \in S$, and

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

Since $J_0 \equiv 0$ is upper semicontinuous and bounded above, so is $J_k^* = T^k(J_0)$, $k = 1, 2, \ldots, N$. The following proposition is obtained by using these facts to parallel the proof of the (F^-) part of Proposition 8.3.

Proposition 8.7 Consider the upper semicontinuous finite horizon model of Definition 8.8. For $k = 1, 2, \ldots, N$, the k-stage optimal cost function J_k^* is upper semicontinuous and bounded above, and $J_k^* = T^k(J_0)$. For each $\varepsilon > 0$, there exists a Borel-measurable, nonrandomized, semi-Markov, ε-optimal policy and a Borel-measurable, (randomized) Markov, ε-optimal policy.

Actually, it is not necessary that S and C be Borel spaces for Proposition 8.7 to hold. Assuming only that S and C are separable metrizable spaces, one can use the results on upper semicontinuity of Section 7.5 and the other assumptions of the upper semicontinuous model to prove the conclusion of Proposition 8.7.

It is not possible to parallel the proof of Proposition 8.4 to show for the upper semicontinuous model that given a sequence of positive numbers $\{\varepsilon_n\}$ with $\varepsilon_n \downarrow 0$, a sequence of Borel-measurable, nonrandomized, Markov policies exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality exists. The set $A(J_{N-1}^*)$ defined by (33) may not be open, so the proof breaks down when one is restricted to Borel-measurable policies.

We conclude this section by pointing out one important case when the disturbance kernel $p(dw|x, u)$ is continuous. If W is n-dimensional Euclidean space and the distribution of w is given by a density $d(w|x, u)$ which is jointly continuous in (x, u) for fixed w, then $p(dw|x, u)$ is continuous. To see this, let G be an open set in W and let $(x_n, u_n) \to (x, u)$ in SC. Then

$$\begin{aligned} \liminf_{k \to \infty} p(G|x_k, u_k) &= \liminf_{k \to \infty} \int_G d(w|x_k, u_k)\, dw \\ &\geq \int_G d(w|x, u)\, dw = p(G|x, u) \end{aligned}$$

by Fatou's lemma. The continuity of $p(dw|x, u)$ follows from Proposition 7.21.†

† Note that by the same argument,

$$\liminf_{k \to \infty} p(G^c|x_k, u_k) \geq p(G^c|x, u),$$

so $p(G|x_k, u_k) \to p(G|x, u)$. Under this condition, the assumption that the system function is continuous in the state (Definitions 8.7(d) and 8.8(c)) can be weakened. See [H3] and [S5].

In fact, it is not necessary that d be continuous in (x, u) for each w, but only that $(x_n, u_n) \to (x, u)$ imply $d(w|x_n, u_n) \to d(w|x, u)$ for Lebesgue almost all w. For example, if $W = R$, the exponential density

$$d(w|x, u) = \begin{cases} \exp[-(w - m(x, u))] & \text{if } w \geq m(x, u), \\ 0 & \text{if } w < m(x, u), \end{cases}$$

where $m: SC \to R$ is continuous, has this property, but need not be continuous in (x, u) for any $w \in R$.

Chapter 9

The Infinite Horizon Borel Models

A first approach to the analysis of the infinite horizon decision model is to treat it as the limit of the finite horizon model as the horizon tends to infinity. In the case (N) of a nonpositive cost per stage and the case (D) of bounded cost per stage and discount factor less than one, this procedure has merit. However, in the case (P) of nonnegative cost per stage, the finite horizon optimal cost functions can fail to converge to the infinite horizon optimal cost function (Example 1 in this chapter), and this failure to converge can occur in such a way that each finite horizon optimal cost function is Borel-measurable, while the infinite horizon optimal cost function is not (Example 2). We thus must develop an independent line of analysis for the infinite horizon model. Our strategy is to define two models, a stochastic one and its deterministic equivalent. There are no measurability restrictions on policies in the deterministic model, and the theory of Part I or of Bertsekas [B4], hereafter abbreviated DPSC, can be applied to it directly. We then transfer this theory to the stochastic model. Sections 9.1–9.3 set up the two models and establish their relationship. Sections 9.4–9.6 analyze the stochastic model via its deterministic counterpart.

9.1 The Stochastic Model

Definition 9.1 An *infinite horizon stochastic optimal control model*, denoted by (SM), is an eight-tuple $(S, C, U, W, p, f, \alpha, g)$ as described in

Definition 8.1. We consider three cases, where Γ is defined by (1) of Chapter 8:

(P) $0 \leq g(x,u)$ for every $(x,u) \in \Gamma$.
(N) $g(x,u) \leq 0$ for every $(x,u) \in \Gamma$.
(D) $0 < \alpha < 1$, and for some $b \in R$, $-b \leq g(x,u) \leq b$ for every $(x,u) \in \Gamma$.

Thus we are really treating three models: (P), (N), and (D). If a result is applicable to one of these models, the corresponding symbol will appear. The assumptions (P), (N), and (D) replace the (F^+) and (F^-) conditions of Chapter 8.

Definition 9.2 A *policy* for (SM) is a sequence $\pi = (\mu_0, \mu_1, \ldots)$ such that for each k, $\mu_k(du_k | x_0, u_0, \ldots, u_{k-1}, x_k)$ is a universally measurable stochastic kernel on C given $SC \cdots CS$ satisfying

$$\mu_k(U(x_k) | x_0, u_0, \ldots, u_{k-1}, x_k) = 1$$

for every $(x_0, u_0, \ldots, u_{k-1}, x_k)$. The concepts of *semi-Markov*, *Markov*, *non-randomized*, and *$\mathcal{F}$-measurable policies* are the same as in Definition 8.2. *We denote by Π' the set of all policies for (SM) and by Π the set of all Markov policies.* If π is a Markov policy of the form $\pi = (\mu, \mu, \ldots)$, it is said to be *stationary*.

As in Chapter 8, we often index S and C for clarity, understanding S_k to be a copy of S and C_k to be a copy of C. Suppose $p \in P(S)$ and $\pi = (\mu_0, \mu_1, \ldots)$ is a policy for (SM). By Proposition 7.45, there is a sequence of unique probability measures $r_N(\pi, p)$ on $S_0 C_0 \cdots S_{N-1} C_{N-1}$, $N = 1, 2, \ldots$, such that for any N and any universally measurable function $h: S_0 C_0 \cdots S_{N-1} C_{N-1} \to R^*$ which satisfies either $\int h^+ \, dr_N(\pi, p) < \infty$ or $\int h^- \, dr_N(\pi, p) < \infty$, (4) of Chapter 8 is satisfied. Furthermore, there exists a unique probability measure $r(\pi, p)$ on $S_0 C_0 S_1 C_1 \cdots$ such that for each N the marginal of $r(\pi, p)$ on $S_0 C_0 \cdots S_{N-1} C_{N-1}$ is $r_N(\pi, p)$. With $r_N(\pi, p)$ and $r(\pi, p)$ determined in this manner, we are ready to define the cost corresponding to a policy.

Definition 9.3 Suppose π is a policy for (SM). The (infinite horizon) *cost corresponding to* π at $x \in S$ is

$$J_\pi(x) = \int \left[\sum_{k=0}^{\infty} \alpha^k g(x_k, u_k) \right] dr(\pi, p_x)$$

$$= \sum_{k=0}^{\infty} \alpha^k \int g(x_k, u_k) \, dr_k(\pi, p_x).^\dagger \tag{1}$$

† The interchange of integration and summation is justified by appeal to the monotone convergence theorem under (P) and (N), and the bounded convergence theorem under (D).

If $\pi = (\mu, \mu, \ldots)$ is stationary, we sometimes write J_μ in place of J_π. The (infinite horizon) *optimal cost* at x is

$$J^*(x) = \inf_{\pi \in \Pi'} J_\pi(x). \tag{2}$$

If $\varepsilon > 0$, the policy π is *ε-optimal at x* provided

$$J_\pi(x) \leq \begin{cases} J^*(x) + \varepsilon & \text{if } J^*(x) > -\infty, \\ -1/\varepsilon & \text{if } J^*(x) = -\infty. \end{cases}$$

If $J_\pi(x) = J^*(x)$, then π is *optimal at x*. If π is ε-optimal or optimal at every $x \in S$, it is said to be *ε-optimal* or *optimal*, respectively.

It is easy to see, using Propositions 7.45 and 7.46, that, for any policy π, $J_\pi(x)$ is universally measurable in x. In fact, if $\pi = (\mu_0, \mu_1, \ldots)$ and $\pi^k = (\mu_0, \ldots, \mu_{k-1})$, then $J_{k,\pi^k}(x)$ defined by (5) of Chapter 8 is universally measurable in x and

$$\lim_{k \to \infty} J_{k,\pi^k}(x) = J_\pi(x) \qquad \forall x \in S. \tag{3}$$

If π is Markov, then (3) can be rewritten in terms of the operators T_{μ_k} of Definition 8.4 as

$$\lim_{k \to \infty} (T_{\mu_0} \cdots T_{\mu_{k-1}})(J_0)(x) = J_\pi(x) \qquad \forall x \in S, \tag{4}$$

which is the infinite horizon analog of Lemma 8.1. If π is a Borel-measurable policy and g is Borel-measurable, then $J_\pi(x)$ is Borel-measurable in x (Proposition 7.29).

It may occur under (P), however, that $\lim_{k \to \infty} J_k^*(x) \neq J^*(x)$, where $J_k^*(x)$ is the optimal k-stage cost defined by (6) of Chapter 8. We offer an example of this.

EXAMPLE 1 Let $S = \{0, 1, 2, \ldots\}$, $C = \{1, 2, \ldots\}$, $U(x) = C$ for every $x \in S$, $\alpha = 1$,

$$f(x, u) = \begin{cases} u & \text{if } x = 0, \\ x - 1 & \text{if } x \neq 0, \end{cases} \qquad g(x, u) = \begin{cases} 1 & \text{if } x = 1, \\ 0 & \text{if } x \neq 1. \end{cases}$$

The problem is deterministic, so the choice of W and $p(dw|x, u)$ is irrelevant. Beginning at $x_0 = 0$, the system moves to some positive integer u_0 at no cost. It then successively moves to $u_0 - 1, u_0 - 2, \ldots$, until it returns to zero and the process begins again. The only transition which incurs a nonzero cost is the transition from one to zero. If the horizon k is finite and u_0 is chosen larger than k, then no cost is incurred before termination, so $J_k^*(0) = 0$. Over the infinite horizon, the transition from one to zero will be made infinitely often, regardless of the policy employed, so $J^*(0) = \infty$.

For $\pi = (\mu_0, \mu_1, \ldots) \in \Pi'$ and $p \in P(S)$, let $q_k(\pi, p)$ be the marginal of $r(\pi, p)$ on $S_k C_k$, $k = 0,1, \ldots$. Then (7) of Chapter 8 holds, and if π is Markov, (8) holds as well. Furthermore, from (1) we have

$$J_\pi(x) = \sum_{k=0}^{\infty} \alpha^k \int_{S_k C_k} g \, dq_k(\pi, p_x) \qquad \forall x \in S, \tag{5}$$

which is the infinite horizon analog of (9) of Chapter 8. Using these facts to parallel the proof of Proposition 8.1, we obtain the following infinite horizon version.

Proposition 9.1 (P)(N)(D) If $x \in S$ and $\pi' \in \Pi'$, then there is a Markov policy π such that

$$J_\pi(x) = J_{\pi'}(x).$$

Corollary 9.1.1 (P)(N)(D) We have

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

where Π is the set of all Markov policies for (SM).

9.2 The Deterministic Model

Definition 9.4 Let $(S, C, U, W, p, f, \alpha, g)$ be an infinite horizon stochastic optimal control model as given by Definition 9.1. The corresponding *infinite horizon deterministic optimal control model*, denoted by (DM), consists of the following:

$P(S)$ *State space.*

$P(SC)$ *Control space.*

$\bar{U}$ *Control constraint.* A function from $P(S)$ to the set of nonempty subsets of $P(SC)$ defined for each $p \in P(S)$ by

$$\bar{U}(p) = \{q \in P(SC) | q(\Gamma) = 1 \text{ and the marginal of } q \text{ on } S \text{ is } p\}, \tag{6}$$

where Γ is given by (1) of Chapter 8.

$\bar{f}$ *System function.* The function from $P(SC)$ to $P(S)$ defined by

$$\bar{f}(q)(\underline{S}) = \int_{SC} t(\underline{S}|x, u) q(d(x, u)) \qquad \forall \underline{S} \in \mathscr{B}_S, \tag{7}$$

where $t(dx'|x, u)$ is given by (3) of Chapter 8.

α *Discount factor.*

$\bar{g}$ *One-stage cost function.* The function from $P(SC)$ to R^* given by

$$\bar{g}(q) = \int_{SC} g(x, u) q(d(x, u)). \tag{8}$$

The model (DM) inherits considerable regularity from (SM). Its state and control spaces $P(S)$ and $P(SC)$ are Borel spaces (Corollary 7.25.1). The system function $\bar{f}$ is Borel-measurable (Proposition 7.26 and Corollary 7.29.1), and the one-stage cost function $\bar{g}$ is lower semianalytic (Corollary 7.48.1). Furthermore, under assumption (P) in (SM), we have $\bar{g} \geq 0$, while under (N), $\bar{g} \leq 0$, and under (D), $0 < \alpha < 1$ and $-b \leq \bar{g} \leq b$.

Definition 9.5 A *policy* for (DM) is a sequence of mappings $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots)$ such that for each k, $\bar{\mu}_k : P(S) \to P(SC)$ and $\bar{\mu}_k(p) \in \bar{U}(p)$ for every $p \in P(S)$. The set of all policies in (DM) will be denoted by $\bar{\Pi}$. We place no measurability requirements on these mappings. A policy $\bar{\pi}$ of the form $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ is said to be *stationary*.

Definition 9.6 Given $p_0 \in P(S)$ and a policy $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots)$ for (DM), the *cost corresponding to* $\bar{\pi}$ at p_0 is

$$\bar{J}_{\bar{\pi}}(p_0) = \sum_{k=0}^{\infty} \alpha^k \bar{g}(q_k), \tag{9}$$

where the control sequence $\{q_k\}$ is generated recursively by means of the equation

$$q_k = \bar{\mu}_k(p_k), \qquad k = 0,1, \ldots, \tag{10}$$

and the system equation

$$p_{k+1} = \bar{f}(q_k), \qquad k = 0, 1, \ldots. \tag{11}$$

If $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ is stationary, we write $\bar{J}_{\bar{\mu}}$ in place of $\bar{J}_{\bar{\pi}}$. The *optimal cost* at p_0 is

$$\bar{J}^*(p_0) = \inf_{\bar{\pi} \in \bar{\Pi}} \bar{J}_{\bar{\pi}}(p_0).$$

The concepts of ε-*optimal* and *optimal policies* for (DM) are the same as those given in Definition 9.3 for (SM).

Definition 9.7 A sequence $(p_0, q_0, q_1, \ldots) \in P(S)P(SC)P(SC)\cdots$ is *admissible in (DM)* if $q_0 \in \bar{U}(p_0)$ and $q_{k+1} \in \bar{U}[\bar{f}(q_k)]$, $k = 0, 1, \ldots$. The set of all admissible sequences will be denoted by Δ.

The admissible sequences are just the sequences of controls $q_0, q_1, \ldots$ together with the initial state p_0 which can be generated by some policy for (DM) via (10) and (11). Except for p_0, the measures p_k are not included in the sequence, but can be recovered as the marginals of the measures q_k on S [cf. (6)].

Definition 9.8 Let $\bar{J} : P(S) \to R^*$ be given and let $\bar{\mu} : P(S) \to P(SC)$ be such that $\bar{\mu}(p) \in \bar{U}(p)$ for every $p \in P(S)$. The *operator* $\bar{T}_{\bar{\mu}}$ mapping $\bar{J}$ into

$\bar{T}_{\bar{\mu}}(\bar{J}):P(S) \to R^*$ is defined by

$$\bar{T}_{\bar{\mu}}(\bar{J})(p) = \bar{g}[\bar{\mu}(p)] + \alpha \bar{J}[\bar{f}(\bar{\mu}(p))] \qquad \forall p \in P(S).$$

The *operator* $\bar{T}$ mapping $\bar{J}$ into $\bar{T}(\bar{J}):P(S) \to R^*$ is defined by

$$\bar{T}(\bar{J})(p) = \inf_{q \in \bar{U}(p)} \{\bar{g}(q) + \alpha \bar{J}[\bar{f}(q)]\} \qquad \forall p \in P(S).$$

Because (DM) is deterministic, it can be studied using results from Part I, Chapters 4 and 5 or from DPSC. This is because there is no need to place measurability restrictions on policies in a deterministic model. The operators $\bar{T}_{\mu}$ and $\bar{T}$ of Definition 9.8 are special cases of those defined in Section 2.1. In the present case, we take $H(p, q, \bar{J})$ to be

$$H(p, q, \bar{J}) = \bar{g}(q) + \alpha \bar{J}[\bar{f}(q)].$$

The monotonicity assumption of Section 2.1 is satisfied by this choice of H. The cost corresponding to a policy $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots)$ as given by (9) is easily seen to be of the form (cf. Section 2.2)

$$\bar{J}_{\bar{\pi}} = \lim_{N \to \infty} (\bar{T}_{\bar{\mu}_0} \cdots \bar{T}_{\bar{\mu}_{N-1}})(\bar{J}_0),$$

where $\bar{J}_0(p) = 0$ for every $p \in P(S)$. It is a straightforward matter to verify that under (D) the contraction assumption of Section 4.1 is satisfied when $\bar{B}$ is taken to be the set of bounded real-valued functions on $P(S)$, m is taken to be one, and $\rho = \alpha$. Under (P), Assumptions I, I.1, and I.2 of Section 5.1 are satisfied, while under (N), Assumptions D, D.1, and D.2 of the same section are in force.

9.3 Relations between the Models

Definition 9.9 Let $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$ be a Markov policy for (SM) and $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots) \in \bar{\Pi}$ a policy for (DM). Let $p_0 \in P(S)$ be given. If for all k

$$\int_{\underline{S}} \mu_k(\underline{C}|x) p_k(dx) = \bar{\mu}_k(p_k)(\underline{S}\underline{C}) \qquad \forall \underline{S} \in \mathscr{B}_S, \quad \underline{C} \in \mathscr{B}_C, \tag{12}$$

where p_k is generated from p_0 by $\bar{\pi}$ via (10) and (11), then π and $\bar{\pi}$ are said to *correspond at* p_0. If π and $\bar{\pi}$ correspond at every $p \in P(S)$, then π and $\bar{\pi}$ are said to *correspond.*

If π and $\bar{\pi}$ correspond at p_0, then the sequence of measures $[q_0(\pi, p_0), q_1(\pi, p_0), \ldots]$ generated from p_0 by π via (8) of Chapter 8 is the same as the sequence $(q_0, q_1, \ldots)$ generated from p_0 by $\bar{\pi}$ via (10) and (11). If π and $\bar{\pi}$ correspond, then they generate the same sequence $(q_0, q_1, \ldots)$ for any initial p_0.

Proposition 9.2 (P)(N)(D) Given a Markov policy $\pi \in \Pi$, there is a corresponding $\bar{\pi} \in \bar{\Pi}$. If $\bar{\pi} \in \bar{\Pi}$ and $p_0 \in P(S)$ are given, then there is a Markov policy $\pi \in \Pi$ corresponding to $\bar{\pi}$ at p_0.

Proof If $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$ is given, then for each k and any $p_k \in P(S)$, there is a unique probability measure on SC, which we denote by $\bar{\mu}_k(p_k)$, satisfying (12) (Proposition 7.45). Furthermore,

$$\bar{\mu}_k(p_k)(\Gamma) = \int_S \mu_k(U(x)|x)p_k(dx) = 1, \tag{13}$$

so $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots)$ is in $\bar{\Pi}$ and corresponds to π. If $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots) \in \bar{\Pi}$ and $p_0 \in P(S)$ are given, let $(p_0, p_1, p_2, \ldots)$ be generated from p_0 by $\bar{\pi}$ via (10) and (11). For each k, choose a Borel-measurable stochastic kernel $\mu_k(du|x)$ which satisfies (12) for this particular p_k (Corollary 7.27.2). Then (13) holds, so

$$\mu_k(U(x)|x) = 1 \tag{14}$$

for p_k almost every x. Altering $\mu_k(du|x)$ on a set of p_k-measure zero if necessary, we may assume that (14) holds for every $x \in S$ and (12) is satisfied. Then $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$ corresponds to $\bar{\pi}$ at p_0. Q.E.D.

Proposition 9.3 (P)(N)(D) Let $p \in P(S)$, $\pi \in \Pi$, and $\bar{\pi} \in \bar{\Pi}$ be given. If π and $\bar{\pi}$ correspond at p, then

$$\bar{J}_{\bar{\pi}}(p) = \int J_\pi(x)p(dx).$$

Proof We have from (7) of Chapter 8, (5), (8), (9), and the monotone or bounded convergence theorems

$$\begin{aligned}
\int J_\pi(x)p(dx) &= \int_S \left[\sum_{k=0}^{\infty} \alpha^k \int_{S_k C_k} g\, dq_k(\pi, p_x) \right] p(dx) \\
&= \sum_{k=0}^{\infty} \alpha^k \int_S \int_{S_k C_k} g\, dq_k(\pi, p_x) p(dx) \\
&= \sum_{k=0}^{\infty} \alpha^k \int_{S_k C_k} g\, dq_k(\pi, p) \\
&= \sum_{k=0}^{\infty} \alpha^k \bar{g}[q_k(\pi, p)] \\
&= \bar{J}_{\bar{\pi}}(p). \qquad \text{Q.E.D.}
\end{aligned}$$

Corollary 9.3.1 (P)(N)(D) Let $x \in S$, $\pi \in \Pi$, and $\bar{\pi} \in \bar{\Pi}$ be given. If π and $\bar{\pi}$ correspond at p_x, then

$$\bar{J}_{\bar{\pi}}(p_x) = J_\pi(x).$$

Corollary 9.3.2 (P)(N)(D) For every $x \in S$,

$$\bar{J}^*(p_x) = J^*(x).$$

Proof Corollaries 9.1.1, 9.3.1, and Proposition 9.2 imply that, for every $x \in S$,

$$\bar{J}^*(p_x) = \inf_{\bar{\pi} \in \bar{\Pi}} \bar{J}_{\bar{\pi}}(p_x) = \inf_{\pi \in \Pi} J_\pi(x) = J^*(x). \qquad \text{Q.E.D.}$$

Corollary 9.3.2 shows that J^* and $\bar{J}^*$ are related, but in a rather weak way that involves $\bar{J}^*$ only on $\bar{S} = \{p_x \in P(S) | x \in S\}$. In Proposition 9.5 we strengthen this relationship, but in order to state that proposition we must show a measurability property of J^*. This is the subject of Proposition 9.4, which we prove with the aid of the following lemma.

Lemma 9.1 The set Δ of admissible sequences in (DM) is an analytic subset of $P(S)P(SC)P(SC)\cdots$.

Proof The set Δ is equal to $A_0 \cap [\bigcap_{k=0}^{\infty} B_k]$, where

$$A_0 = \{(p_0, q_0, q_1, \ldots) | q_0 \in \bar{U}(p_0)\},$$
$$B_k = \{(p_0, q_0, q_1, \ldots) | q_{k+1} \in \bar{U}[\bar{f}(q_k)]\}.$$

By Corollary 7.35.2, it suffices to show that A_0 and B_k, $k = 0, 1, \ldots$, are analytic. Using the result of Proposition 7.38, this will follow if we show that

$$A = \{(p_1, q_1) \in P(S)P(SC) | q_1 \in \bar{U}(p_1)\},$$
$$B = \{(q_0, q_1) \in P(SC)P(SC) | q_1 \in \bar{U}[\bar{f}(q_0)]\}$$

are analytic. Let $P(\Gamma) = \{q \in P(SC) | q(\Gamma) = 1\}$, where Γ is given by (1) of Chapter 8. Then $P(\Gamma)$ is analytic (Proposition 7.43). Equation (6) implies that A is the intersection of the analytic set $P(S)P(\Gamma)$ (Proposition 7.38) with the graph of the function $\sigma: P(SC) \to P(S)$ which maps q into its marginal on S. It is easily verified that σ is continuous (Proposition 7.21(a) and (b)), so $\mathrm{Gr}(\sigma)$ is Borel (Corollary 7.14.1). Therefore, A is analytic. The set B is the inverse image of A under the Borel-measurable mapping $(q_0, q_1) \to [\bar{f}(q_0), q_1]$, so is also analytic (Proposition 7.40). Q.E.D.

Proposition 9.4 (P)(N)(D) The function $\bar{J}^*: P(S) \to R^*$ is lower semi-analytic.

Proof Define $G: \Delta \to R^*$ by

$$G(p_0, q_0, q_1, \ldots) = \sum_{k=0}^{\infty} \alpha^k \bar{g}(q_k), \tag{15}$$

where Δ is the set of admissible sequences (Definition 9.7). Then G is lower semianalytic by Lemma 7.30(2), (4) and Lemma 9.1. By the definition of $\bar{J}^*$ and Δ, we have

$$\bar{J}^*(p_0) = \inf_{(q_0,q_1,\ldots)\in\Delta_{p_0}} G(p_0, q_0, q_1, \ldots) \qquad \forall p_0 \in P(S), \tag{16}$$

so $\bar{J}^*$ is lower semianalytic by Proposition 7.47. Q.E.D.

Corollary 9.4.1 (P)(N)(D) The function $J^*: S \to R^*$ is lower semianalytic.

Proof By Corollary 9.3.2,

$$J^*(x) = \bar{J}^*[\delta(x)] \qquad \forall x \in S,$$

where $\delta(x) = p_x$ is the homeomorphism defined in Corollary 7.21.1. Apply Lemma 7.30(3) and Proposition 9.4 to conclude that J^* is lower semianalytic. Q.E.D.

Lemma 9.2 Given $p \in P(S)$ and $\varepsilon > 0$, there exists a policy $\bar{\pi}$ for (DM) such that

(P)(D) $$\bar{J}_{\bar{\pi}}(p) \le \int J^*(x)p(dx) + \varepsilon,$$

(N) $$\bar{J}_{\bar{\pi}}(p) \le \begin{cases} \int J^*(x)p(dx) + \varepsilon & \text{if } \int J^*(x)p(dx) > -\infty, \\ -1/\varepsilon & \text{if } \int J^*(x)p(dx) = -\infty. \end{cases}$$

Proof As a consequence of Corollary 9.4.1, $\int J^*(x)p(dx)$ is well defined. Let $p \in P(S)$ and $\varepsilon > 0$ be given. Let $G: \Delta \to R^*$ be defined by (15). Proposition 7.50 guarantees that under (P) and (D) there exists a universally measurable selector $\varphi: P(S) \to P(SC)P(SC)\cdots$ such that $(p, \varphi(p)) \in \Delta$ for every $p \in P(S)$ and

$$G[p, \varphi(p)] \le \bar{J}^*(p) + \varepsilon \qquad \forall p \in P(S).$$

Let $\sigma: S \to P(SC)P(SC)\cdots$ be defined by $\sigma(x) = \varphi(p_x)$. Then σ is universally measurable (Proposition 7.44) and

$$G[p_x, \sigma(x)] \le J^*(x) + \varepsilon \qquad \forall x \in S. \tag{17}$$

Under (N), there exists a universally measurable $\sigma: S \to P(SC)P(SC)\cdots$ such that for every $x \in S$, $(p_x, \sigma(x)) \in \Delta$ and

$$G[p_x, \sigma(x)] \le \begin{cases} J^*(x) + \varepsilon & \text{if } J^*(x) > -\infty, \\ -(1 + \varepsilon^2)/\varepsilon p_\infty(J^*) & \text{if } J^*(x) = -\infty, \end{cases} \tag{18}$$

where $p_\infty(J^*) = p(\{x | J^*(x) = -\infty\})$ if $p(\{x | J^*(x) = -\infty\}) > 0$ and $p_\infty(J^*) = 1$ otherwise.

Denote $\sigma(x) = [q_0(d(x_0, u_0)|x), q_1(d(x_1, u_1)|x), \ldots]$. Each $q_k(d(x_k, u_k)|x)$ is a universally measurable stochastic kernel on $S_k C_k$ given S. Furthermore,

$$q_0(d(x_0, u_0)|x) \in \bar{U}(p_x) \qquad \forall x \in S,$$

and, for $k = 0, 1, \ldots,$

$$q_{k+1}(d(x_{k+1}, u_{k+1})|x) \in \bar{U}(\bar{f}[q_k(d(x_k, u_k)|x)]) \qquad \forall x \in S.$$

For $k = 0, 1, \ldots,$ define $\bar{q}_k \in P(SC)$ by

$$\bar{q}_k(B) = \int q_k(B|x)p(dx) \qquad \forall B \in \mathscr{B}_{SC}.$$

Then $\bar{q}_k(\Gamma) = 1$, $k = 0, 1, \ldots$. We show that $(p, \bar{q}_0, \bar{q}_1, \ldots) \in \Delta$. Since the marginal of $q_0(d(x_0, u_0)|x)$ on S_0 is p_x, we have

$$\bar{q}_0(\underline{S}_0 C_0) = \int_S q_0(\underline{S}_0 C_0|x)p(dx) = \int_S \chi_{\underline{S}_0}(x)p(dx) = p(\underline{S}_0) \qquad \forall \underline{S}_0 \in \mathscr{B}_S,$$

so $\bar{q}_0 \in \bar{U}(p)$. For $k = 0, 1, \ldots,$ we have

$$\begin{aligned}
\bar{q}_{k+1}(\underline{S}_{k+1} C_{k+1}) &= \int_S q_{k+1}(\underline{S}_{k+1} C_{k+1}|x)p(dx) \\
&= \int_S \int_{S_k C_k} t(\underline{S}_{k+1}|x_k, u_k) q_k(d(x_k, u_k)|x)p(dx) \\
&= \int_{S_k C_k} t(\underline{S}_{k+1}|x_k, u_k)\bar{q}_k(d(x_k, u_k)) \qquad \forall \underline{S}_{k+1} \in \mathscr{B}_S.
\end{aligned}$$

Therefore $\bar{q}_{k+1} \in \bar{U}[\bar{f}(\bar{q}_k)]$ and $(p, \bar{q}_0, \bar{q}_1, \ldots) \in \Delta$.

Let $\bar{\pi}$ be any policy for (DM) which generates the admissible sequence $(p, \bar{q}_0, \bar{q}_1, \ldots)$. Then under (P) and (D), we have from (17) and the monotone or bounded convergence theorem

$$\begin{aligned}
\bar{J}_{\bar{\pi}}(p) &= G(p, \bar{q}_0, \bar{q}_1, \ldots) \\
&= \sum_{k=0}^{\infty} \alpha^k \int_{S_k C_k} g(x_k, u_k)\bar{q}_k(d(x_k, u_k)) \\
&= \sum_{k=0}^{\infty} \alpha^k \int_S \int_{S_k C_k} g(x_k, u_k) q_k(d(x_k, u_k)|x)p(dx) \\
&= \int_S \left[\sum_{k=0}^{\infty} \alpha^k \int_{S_k C_k} g(x_k, u_k) q_k(d(x_k, u_k)|x) \right] p(dx) \\
&= \int_S G[p_x, \sigma(x)]p(dx) \\
&\leq \int_S J^*(x)p(dx) + \varepsilon.
\end{aligned}$$

Under (N), we have from the monotone convergence theorem

$$\bar{J}_{\bar{\pi}}(p) = \int_S G[p_x, \sigma(x)]p(dx). \tag{19}$$

If $p(\{x|J^*(x) = -\infty\}) = 0$, (18) and (19) imply

$$\bar{J}_{\bar{\pi}}(p) \leq \int J^*(x)p(dx) + \varepsilon,$$

where both sides may be $-\infty$. If $p(\{x|J^*(x) = -\infty\}) > 0$, then $\int J^*(x)$ $p(dx) = -\infty$ and we have, from (18) and (19),

$$\begin{aligned}\bar{J}_{\bar{\pi}}(p) &\leq \int_{\{x|J^*(x) > -\infty\}} [J^*(x) + \varepsilon]p(dx) - (1 + \varepsilon^2)/\varepsilon \\ &\leq \varepsilon - (1 + \varepsilon^2)/\varepsilon = -1/\varepsilon. \qquad \text{Q.E.D.}\end{aligned}$$

Proposition 9.5 (P)(N)(D) For every $p \in P(S)$,

$$\bar{J}^*(p) = \int J^*(x)p(dx).$$

Proof Lemma 9.2 shows that

$$\bar{J}^*(p) \leq \int J^*(x)p(dx) \qquad \forall p \in P(S).$$

For the reverse inequality, let p be in $P(S)$ and let $\bar{\pi}$ be a policy for (DM). There exists a policy $\pi \in \Pi$ corresponding to $\bar{\pi}$ at p (Proposition 9.2), and, by Proposition 9.3,

$$\bar{J}_{\bar{\pi}}(p) = \int J_\pi(x)p(dx) \geq \int J^*(x)p(dx).$$

By taking the infimum of the left-hand side over $\bar{\pi} \in \bar{\Pi}$, we obtain the desired result. Q.E.D.

Propositions 9.3 and 9.5 are the key relationships between (SM) and (DM). As an example of their implications, consider the following corollary.

Corollary 9.5.1 (P)(N)(D) Suppose $\pi \in \Pi$ and $\bar{\pi} \in \bar{\Pi}$ are corresponding policies for (SM) and (DM). Then π is optimal if and only if $\bar{\pi}$ is optimal.

Proof If π is optimal, then

$$\bar{J}_{\bar{\pi}}(p) = \int J_\pi(x)p(dx) = \int J^*(x)p(dx) = \bar{J}^*(p) \qquad \forall p \in P(S).$$

If $\bar{\pi}$ is optimal, then

$$J_\pi(x) = \bar{J}_{\bar{\pi}}(p_x) = \bar{J}^*(p_x) = J^*(x) \qquad \forall x \in S. \qquad \text{Q.E.D.}$$

The next corollary is a technical result needed for Chapter 10.

Corollary 9.5.2 (P)(N)(D) For every $p \in P(S)$,

$$\int J^*(x)p(dx) = \inf_{\pi \in \Pi} \int J_\pi(x)p(dx).$$

Proof By Propositions 9.2 and 9.3,

$$\bar{J}^*(p) = \inf_{\pi \in \Pi} \int J_\pi(x)p(dx) \qquad \forall p \in P(S).$$

Apply Proposition 9.5. Q.E.D.

We now explore the connections between the operators T_μ and $\bar{T}_{\bar{\mu}}$ and the operators T and $\bar{T}$. The first proposition is a direct consequence of the definitions. We leave the verification to the reader.

Proposition 9.6 (P)(N)(D) Let $J: S \to R^*$ be universally measurable and satisfy $J \geq 0$, $J \leq 0$, or $-c \leq J \leq c$, $c < \infty$, according as (P), (N), or (D) is in force. Let $\bar{J}: P(S) \to R^*$ be defined by

$$\bar{J}(p) = \int J(x)p(dx) \qquad \forall p \in P(S),$$

and suppose $\bar{\mu}: P(S) \to P(SC)$ is of the form

$$\bar{\mu}(p)(\underline{S}\underline{C}) = \int_{\underline{S}} \mu(\underline{C}|x)p(dx) \qquad \forall \underline{S} \in \mathscr{B}_S, \quad \underline{C} \in \mathscr{B}_C$$

for some $\mu \in U(C|S)^\dagger$. Then $\bar{\mu}(p) \in \bar{U}(p)$ for every $p \in P(S)$, and

$$\bar{T}_{\bar{\mu}}(\bar{J})(p) = \int T_\mu(J)(x)p(dx) \qquad \forall p \in P(S).$$

Proposition 9.7 (P)(N)(D) Let $J: S \to R^*$ be lower semianalytic and satisfy $J \geq 0$, $J \leq 0$, or $-c \leq J \leq c$, $c < \infty$, according as (P), (N), or (D) is in force. Let $\bar{J}: P(S) \to R^*$ be defined by

$$\bar{J}(p) = \int J(x)p(dx) \qquad \forall p \in P(S). \tag{20}$$

Then

$$\bar{T}(\bar{J})(p) = \int T(J)(x)p(dx) \qquad \forall p \in P(S).$$

Proof For $p \in P(S)$ and $q \in \bar{U}(p)$ we have

$$\begin{aligned} \bar{g}(q) + \alpha \bar{J}[\bar{f}(q)] &= \int_{SC} [g(x,u) + \alpha \int_S J(x')t(dx'|x,u)]q(d(x,u)) \\ &\geq \int_S T(J)(x)p(dx), \end{aligned}$$

† The set $U(C|S)$, defined in Section 8.2, is the collection of universally measurable stochastic kernels μ on C given S which satisfy $\mu(U(x)|x) = 1$ for every $x \in S$.

which implies

$$\bar{T}(\bar{J})(p) \geq \int T(J)(x)p(dx).$$

Given $\varepsilon > 0$, Lemma 8.2 implies that there exists $\mu \in U(C|S)$ such that

$$\int_C [g(x,u) + \alpha \int_S J(x')t(dx'|x,u)]\mu(du|x) \leq T(J)(x) + \varepsilon.$$

Let $q \in \bar{U}(p)$ be such that

$$q(\underline{S}\underline{C}) = \int_{\underline{S}} \mu(\underline{C}|x)p(dx) \qquad \forall \underline{S} \in \mathscr{B}_S, \underline{C} \in \mathscr{B}_C.$$

Then

$$\begin{aligned}\bar{T}(\bar{J})(p) &\leq \int_{SC} \left[g(x,u) + \alpha \int_S J(x')t(dx'|x,u) \right] q(d(x,u)) \\ &= \int_S \int_C \left[g(x,u) + \alpha \int_S J(x')t(dx'|x,u) \right] \mu(du|x)p(dx) \\ &\leq \int T(J)(x)p(dx) + \varepsilon,\end{aligned}$$

where $\int T(J)(x)p(dx) + \varepsilon$ may be $-\infty$. Therefore,

$$\bar{T}(\bar{J})(p) \leq \int T(J)(x)p(dx). \qquad \text{Q.E.D.}$$

9.4 The Optimality Equation—Characterization of Optimal Policies

As noted following Definition 9.8, the model (DM) is a special case of that considered in Part I and DPSC[†]. This allows us to easily obtain many results for both (SM) and (DM). A prime example of this is the next proposition.

Proposition 9.8 (P)(N)(D) We have

$$\bar{J}^* = \bar{T}(\bar{J}^*), \tag{21}$$

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

Proof The optimality equation (21) for (DM) follows from Propositions 4.2(a), 5.2, and 5.3 or from DPSC, Chapter 6, Proposition 2 and Chapter 7,

[†] Whereas we allow $\bar{g}$ to be extended real-valued, in Chapter 7 of DPSC the one-stage cost function is assumed to be real-valued. This more restrictive assumption is not essential to any of the results we quote from DPSC.

Proposition 1. We have then, for any $x \in S$,

$$J^*(x) = \bar{J}^*(p_x) = \bar{T}(\bar{J}^*)(p_x) = T(J^*)(x)$$

by Propositions 9.5 and 9.7, so (22) holds as well. Q.E.D.

Proposition 9.9 (P)(N)(D) If $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ is a stationary policy for (DM), then $\bar{J}_{\bar{\mu}} = \bar{T}_{\bar{\mu}}(\bar{J}_{\bar{\mu}})$. If $\pi = (\mu, \mu, \ldots)$ is a stationary policy for (SM), then $J_\mu = T_\mu(J_\mu)$.

Proof For (DM) this result follows from Proposition 4.2(b), Corollary 5.2.1, and Corollary 5.3.2 or from DPSC, Chapter 6, Corollary 2.1 and Chapter 7, Corollary 1.1. Let $\pi = (\mu, \mu, \ldots)$ be a stationary policy for (SM) and let $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ be a policy for (DM) corresponding to π. Then for each $x \in S$,

$$J_\mu(x) = \bar{J}_{\bar{\mu}}(p_x) = \bar{T}_{\bar{\mu}}(\bar{J}_{\bar{\mu}})(p_x) = T_\mu(J_\mu)(x)$$

by Propositions 9.3 and 9.6. Q.E.D.

Note that Proposition 9.9 for (SM) cannot be deduced from Proposition 9.8 by considering a modified (SM) with control constraint of the form

$$U_\mu(x) = \{\mu(x)\} \qquad \forall x \in S, \tag{23}$$

as was done in the proof of Corollary 5.2.1. Even if μ is nonrandomized so that (23) makes sense, the set

$$\Gamma_\mu = \{(x, u) | x \in S, u \in U_\mu(x)\}$$

may not be analytic, so U_μ is not an acceptable control constraint.

The optimality equations are necessary conditions for the optimal cost functions, but except in case (D) they are by no means sufficient. We have the following partial sufficiency results.

Proposition 9.10

(P) If $\bar{J}: P(S) \to [0, \infty]$ and $\bar{J} \geq \bar{T}(\bar{J})$, then $\bar{J} \geq \bar{J}^*$.
If $J: S \to [0, \infty]$ is lower semianalytic and $J \geq T(J)$, then $J \geq J^*$.

(N) If $\bar{J}: P(S) \to [-\infty, 0]$ and $\bar{J} \leq \bar{T}(\bar{J})$, then $\bar{J} \leq \bar{J}^*$.
If $J: S \to [-\infty, 0]$ is lower semianalytic and $J \leq T(J)$, then $J \leq J^*$.

(D) If $\bar{J}: P(S) \to [-c, c]$, $c < \infty$, and $\bar{J} = \bar{T}(\bar{J})$, then $\bar{J} = \bar{J}^*$.
If $J: S \to [-c, c]$, $c < \infty$, is lower semianalytic and $J = T(J)$, then $J = J^*$.

Proof We consider first the statements for (DM). The result under (P) follows from Proposition 5.2, the result under (N) from Proposition 5.3, and the result under (D) from Proposition 4.2(a). These results for (DM)

follow from Proposition 2 and trivial modifications of the proof of Proposition 9 of DPSC, Chapter 6.

We now establish the (SM) part of the proposition under (P). Cases (N) and (D) are handled in the same manner. Given a lower semianalytic function $J:S \to [0, \infty]$ satisfying $J \geq T(J)$, define $\bar{J}:P(S) \to [0, \infty]$ by (20). Then

$$\bar{J}(p) = \int J(x)p(dx) \geq \int T(J)(x)p(dx) = \bar{T}(\bar{J})(p) \qquad \forall p \in P(S)$$

by Proposition 9.7. By the result for (DM), $\bar{J} \geq \bar{J}^*$. In particular,

$$J(x) = \bar{J}(p_x) \geq \bar{J}^*(p_x) = J^*(x) \qquad \forall x \in S. \qquad \text{Q.E.D.}$$

Proposition 9.11 Let $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ and $\pi = (\mu, \mu, \ldots)$ be stationary policies in (DM) and (SM), respectively.

(P) If $\bar{J}:P(S) \to [0, \infty]$ and $\bar{J} \geq \bar{T}_{\bar{\mu}}(\bar{J})$, then $\bar{J} \geq \bar{J}_{\bar{\mu}}$.
If $J:S \to [0, \infty]$ is universally measurable and $J \geq T_\mu(J)$, then $J \geq J_\mu$.

(N) If $\bar{J}:P(S) \to [-\infty, 0]$ and $\bar{J} \leq \bar{T}_{\bar{\mu}}(\bar{J})$, then $\bar{J} \leq \bar{J}_{\bar{\mu}}$.
If $J:S \to [-\infty, 0]$ is universally measurable and $J \leq T_\mu(J)$, then $J \leq J_\mu$.

(D) If $\bar{J}:P(S) \to [-c, c]$, $c < \infty$, and $\bar{J} = \bar{T}_{\bar{\mu}}(\bar{J})$, then $\bar{J} = \bar{J}_{\bar{\mu}}$.
If $J:S \to [-c, c]$, $c < \infty$, is universally measurable and $J = T_\mu(J)$, then $J = J_\mu$.

Proof The (DM) results follow from Proposition 4.2(b) and Corollaries 5.2.1 and 5.3.2 or from DPSC, Corollary 2.1 and trivial modifications of Corollary 9.1 of Chapter 6. The (SM) results follow from the (DM) results and Proposition 9.6 in a manner similar to the proof of Proposition 9.10. Q.E.D.

Proposition 9.11 implies that under (P), J_μ is the smallest nonnegative universally measurable solution to the functional equation

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

Under (D), J_μ is the only bounded universally measurable solution to this equation. This provides us with a simple necessary and sufficient condition for a stationary policy to be optimal under (P) and (D).

Proposition 9.12 (P)(D) Let $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ and $\pi = (\mu, \mu, \ldots)$ be stationary policies in (DM) and (SM), respectively. The policy $\bar{\pi}$ is optimal if and only if $\bar{J}^* = \bar{T}_{\bar{\mu}}(\bar{J}^*)$. The policy π is optimal if and only if $J^* = T_\mu(J^*)$.

Proof If $\bar{\pi}$ is optimal, then $\bar{J}_{\bar{\mu}} = \bar{J}^*$. By Proposition 9.9, $\bar{J}^* = \bar{T}_{\bar{\mu}}(\bar{J}^*)$. Conversely, if $\bar{J}^* = \bar{T}_{\bar{\mu}}(\bar{J}^*)$, then, by Proposition 9.11, $\bar{J}^* \geq \bar{J}_{\bar{\mu}}$ and $\bar{\pi}$ is

optimal. The proof for (SM) follows from the (SM) parts of the same propositions. Q.E.D.

Corollary 9.12.1 (P)(D) There is an optimal nonrandomized stationary policy for (SM) if and only if for each $x \in S$ the infimum in

$$\inf_{u \in U(x)} \left\{ g(x,u) + \alpha \int J^*(x')t(dx'|x,u) \right\} \tag{24}$$

is achieved.

Proof If the infimum in (24) is achieved for every $x \in S$, then by Proposition 7.50 there is a universally measurable selector $\mu: S \to C$ whose graph lies in Γ and for which

$$g[x,\mu(x)] + \alpha \int J^*(x')t(dx'|x,\mu(x)) = \inf_{u \in U(x)} \left\{ g(x,u) + \alpha \int J^*(x')t(dx'|x,u) \right\} \qquad \forall x \in S.$$

Then by Proposition 9.8

$$T_\mu(J^*) = T(J^*) = J^*,$$

so $\pi = (\mu, \mu, \ldots)$ is optimal by Proposition 9.12.

If $\pi = (\mu, \mu, \ldots)$ is an optimal nonrandomized stationary policy for (SM), then by Propositions 9.8 and 9.9

$$T_\mu(J^*) = T_\mu(J_\mu) = J_\mu = J^* = T(J^*),$$

so $\mu(x)$ achieves the infimum in (24) for every $x \in S$. Q.E.D.

In Proposition 9.19, we show that under (P) or (D), the existence of any optimal policy at all implies the existence of an optimal policy that is nonrandomized and stationary. This means that Corollary 9.12.1 actually gives a necessary and sufficient condition for the existence of an optimal policy.

Under (N) we can use Proposition 9.10 to obtain a necessary and sufficient condition for a stationary policy to be optimal. This condition is not as useful as that of Proposition 9.12, however, since it cannot be used to construct a stationary optimal policy in the manner of Corollary 9.12.1.

Proposition 9.13 (N)(D) Let $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ and $\pi = (\mu, \mu, \ldots)$ be stationary policies in (DM) and (SM), respectively. The policy $\bar{\pi}$ is optimal if and only if $\bar{J}_{\bar{\mu}} = \bar{T}(\bar{J}_{\bar{\mu}})$. The policy π is optimal if and only if $J_\mu = T(J_\mu)$.

Proof If $\bar{\pi}$ is optimal, then $\bar{J}_{\bar{\mu}} = \bar{J}^*$. By Proposition 9.8

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

Conversely, if $\bar{J}_{\bar{\mu}} = \bar{T}(\bar{J}_{\bar{\mu}})$, then Proposition 9.10 implies that $\bar{J}_{\bar{\mu}} \leq \bar{J}^*$ and $\bar{\pi}$ is optimal.

If π is optimal, $J_\mu = T(J_\mu)$ by the (SM) part of Proposition 9.8. The converse is more difficult, since the (SM) part of Proposition 9.10 cannot be invoked without knowing that J_μ is lower semianalytic. Let $\bar{\pi} = (\bar{\mu}, \bar{\mu}, \ldots)$ be a policy for (DM) corresponding to $\pi = (\mu, \mu, \ldots)$, so that $\bar{J}_{\bar{\mu}}(p) = \int J_\mu(x)p(dx)$ for every $p \in P(S)$. Then for fixed $p \in P(S)$ and $q \in \bar{U}(p)$,

$$\begin{aligned}\bar{g}(q) + \alpha \bar{J}_{\bar{\mu}}[\bar{f}(q)] &= \int_{SC}\left[g(x,u) + \alpha \int_S J_\mu(x')t(dx'|x,u) \right] q(d(x,u)) \\ &\geq \int \inf_{u \in U(x)} \left\{ g(x,u) + \alpha \int_S J_\mu(x')t(dx'|x,u) \right\} p(dx),\end{aligned}$$

provided the integrand

$$T(J_\mu)(x) = \inf_{u \in U(x)} \left\{ g(x,u) + \alpha \int_S J_\mu(x')t(dx'|x,u) \right\}$$

is universally measurable in x. But $T(J_\mu) = J_\mu$ by assumption, which is universally measurable, so

$$\bar{g}(q) + \alpha \bar{J}_{\bar{\mu}}[\bar{f}(q)] \geq \int J_\mu(x)p(dx) = \bar{J}_{\bar{\mu}}(p).$$

By taking the infimum of the left-hand side over $q \in \bar{U}(p)$ and using Proposition 9.9, we see that

$$\bar{T}(\bar{J}_{\bar{\mu}}) \geq \bar{J}_{\bar{\mu}} = \bar{T}_{\bar{\mu}}(\bar{J}_{\bar{\mu}}).$$

The reverse inequality always holds, and by the result already proved for (DM), $\bar{\pi}$ is optimal. The optimality of π follows from Corollary 9.5.1. Q.E.D.

9.5 Convergence of the Dynamic Programming Algorithm—Existence of Stationary Optimal Policies

Definition 9.10 The *dynamic programming algorithm* is defined recursively for (DM) and (SM) by

$$\begin{aligned}\bar{J}_0(p) &= 0 \qquad \forall p \in P(S), \\ \bar{J}_{k+1}(p) &= \bar{T}(\bar{J}_k)(p) \qquad \forall p \in P(S), \quad k = 0, 1, \ldots, \\ J_0(x) &= 0 \qquad \forall x \in S, \\ J_{k+1}(x) &= T(J_k)(x) \qquad \forall x \in S, \quad k = 0, 1, \ldots.\end{aligned}$$

We know from Proposition 8.2 that this algorithm generates the k-stage optimal cost functions J_k^*. For simplicity of notation, we suppress the $*$ here. At present we are concerned with the infinite horizon case and the possibility that J_k may converge to J^* as $k \to \infty$.

Under (P), $\bar{J}_0 \le \bar{J}_1$ and so $\bar{J}_1 = \bar{T}(\bar{J}_0) \le \bar{T}(\bar{J}_1) = \bar{J}_2$. Continuing, we see that $\bar{J}_k$ is an increasing sequence of functions, and so $\bar{J}_\infty = \lim_{k\to\infty} \bar{J}_k$ exists and takes values in $[0, +\infty]$. Under (N), $\bar{J}_k$ is a decreasing sequence of functions and $\bar{J}_\infty$ exists, taking values in $[-\infty, 0]$. Under (D), we have

$$\begin{aligned} \bar{J}_0 &\le b + \bar{T}(\bar{J}_0), \\ 0 \le b + \bar{T}(\bar{J}_0) &\le b + \bar{T}[b + \bar{T}(\bar{J}_0)] = (1+\alpha)b + \bar{T}^2(\bar{J}_0), \\ 0 \le b + \bar{T}[b + \bar{T}(\bar{J}_0)] &= (1+\alpha)b + \bar{T}^2(\bar{J}_0) \le b + \bar{T}[(1+\alpha)b + \bar{T}^2(\bar{J}_0)] \\ &= (1 + \alpha + \alpha^2)b + \bar{T}^3(\bar{J}_0), \end{aligned}$$

and, in general,

$$0 \le b \sum_{j=0}^{k-2} \alpha^j + \bar{T}^{k-1}(\bar{J}_0) \le b \sum_{j=0}^{k-1} \alpha^j + \bar{T}^k(\bar{J}_0).$$

As $k \to \infty$, we see that $b\sum_{j=0}^{k-1} \alpha^j + \bar{T}^k(\bar{J}_0)$ increases to a limit. But $b\sum_{j=0}^{\infty} \alpha^j = b/(1-\alpha)$, so $\bar{J}_\infty = \lim_{k\to\infty} \bar{T}^k(\bar{J}_0)$ exists and satisfies

$$-b/(1-\alpha) \le \bar{J}_\infty.$$

Similarly, we have

$$\bar{J}_\infty \le b/(1-\alpha).$$

Now if $\bar{J}: P(S) \to [-c, c]$, $c < \infty$, then

$$\bar{J}_0 \le \bar{J} + c, \qquad \bar{T}(\bar{J}_0) \le \bar{T}(\bar{J} + c) = \alpha c + \bar{T}(\bar{J}),$$

and, in general,

$$\bar{T}^k(\bar{J}_0) \le \alpha^k c + \bar{T}^k(\bar{J}).$$

It follows that

$$\bar{J}_\infty \le \liminf_{k\to\infty} \bar{T}^k(\bar{J}),$$

and by a similar argument beginning with $\bar{J} - c \le \bar{J}_0$, we can show that $\limsup_{k\to\infty} \bar{T}^k(\bar{J}) \le \bar{J}_\infty$. This shows that under (D), if $\bar{J}$ is any bounded real-valued function on $P(S)$, then $\bar{J}_\infty = \lim_{k\to\infty} \bar{T}^k(\bar{J})$.

The same arguments can be used to establish the existence of $J_\infty = \lim_{k\to\infty} J_k$. Under (P), $J_\infty: S \to [0, +\infty]$; under (N), $J_\infty: S \to [-\infty, 0]$; and under (D), $J_\infty = \lim_{k\to\infty} T^k(J)$ takes values in $[-b/(1-\alpha), b/(1-\alpha)]$ where $J: S \to [-c, c]$, $c < \infty$, is lower semianalytic. Note that in every case, J_∞ is lower semianalytic by Lemma 7.30(2).

Lemma 9.3 (P)(N)(D) For every $p \in P(S)$,

$$\bar{J}_k(p) = \int J_k(x)p(dx), \qquad k = 0, 1, \ldots, \quad k = \infty.$$

Proof For $k = 0, 1, \ldots$, the lemma follows from Proposition 9.7 by induction. When $k = \infty$, the lemma follows from the monotone convergence theorem under (P) and (N) and the bounded convergence theorem under (D). Q.E.D.

Proposition 9.14 (N)(D) We have

$$\bar{J}_\infty = \bar{J}^*, \tag{25}$$

$$J_\infty = J^*. \tag{26}$$

Indeed, under (D) the dynamic programming algorithm can be initiated from any $\bar{J}: P(S) \to [-c, c]$, $c < \infty$, or lower semianalytic $J: S \to [-c, c]$, $c < \infty$, and converges uniformly, i.e.,

$$\lim_{k \to \infty} \sup_{p \in P(S)} |\bar{T}^k(\bar{J})(p) - \bar{J}^*(p)| = 0, \tag{27}$$

$$\lim_{k \to \infty} \sup_{x \in S} |T^k(J)(x) - J^*(x)| = 0. \tag{28}$$

Proof The result for (DM) follows from Proposition 4.2(c) and 5.7 or from DPSC, Chapter 6, Proposition 3 and Chapter 7, Proposition 4. By Lemma 9.3,

$$J_k(x) = \bar{J}_k(p_x) \qquad \forall x \in S, \quad k = 0, 1, \ldots, \quad k = \infty,$$

so (25) implies (26). Under (D), if a lower semianalytic function $J: S \to [-c, c]$, $c < \infty$, is given, then define $\bar{J}: P(S) \to [-c, c]$ by (20). Equation (28) now follows from (27) and Propositions 9.5 and 9.7. Q.E.D.

Case (D) is the best suited for computational procedures. The machinery developed thus far can be applied to Proposition 4.6 or to DPSC, Chapter 6, Proposition 4, to show the validity for (SM) of the error bounds given there. We provide the theorem for (SM). The analogous result is of course true for (DM).

Proposition 9.15 (D) Let $J: S \to [-c, c]$, $c < \infty$, be lower semianalytic. Then for all $x \in S$ and $k = 0, 1, \ldots$,

$$\begin{aligned} T^k(J)(x) + b_k &\le T^{k+1}(J)(x) + b_{k+1} \\ &\le J^*(x) \le T^{k+1}(J)(x) + \bar{b}_{k+1} \le T^k(J)(x) + \bar{b}_k, \end{aligned} \tag{29}$$

where

$$b_k = [\alpha/(1 - \alpha)] \inf_{x \in S} [T^k(J)(x) - T^{k-1}(J)(x)], \tag{30}$$

$$\bar{b}_k = [\alpha/(1 - \alpha)] \sup_{x \in S} [T^k(J)(x) - T^{k-1}(J)(x)]. \tag{31}$$

Proof Given a lower semianalytic function $J: S \to [-c, c]$, $c < \infty$, define $\bar{J}: P(S) \to [-c, c]$ by (20). By Proposition 9.7,

$$\bar{T}^k(\bar{J})(p) = \int T^k(J)(x)p(dx) \qquad \forall p \in P(S), \quad k = 0, 1, \ldots .$$

Therefore

$$b_k = [\alpha/(1-\alpha)] \inf_{p \in P(S)} [\bar{T}^k(\bar{J})(p) - \bar{T}^{k-1}(\bar{J})(p)],$$

$$\bar{b}_k = [\alpha/(1-\alpha)] \sup_{p \in P(S)} [\bar{T}^k(\bar{J})(p) - \bar{T}^{k-1}(\bar{J})(p)],$$

where b_k and $\bar{b}_k$ are defined by (30) and (31). Taking $\alpha_1 = \alpha_2 = \alpha$ in Proposition 4.6 or using the proof of Proposition 4, Chapter 6 of DPSC, we obtain

$$\begin{aligned} \bar{T}^k(\bar{J})(p) + b_k &\le \bar{T}^{k+1}(\bar{J})(p) + b_{k+1} \\ &\le \bar{J}^*(p) \le \bar{T}^{k+1}(\bar{J})(p) + \bar{b}_{k+1} \le \bar{T}^k(\bar{J})(p) + \bar{b}_k. \end{aligned}$$

Substituting $p = p_x$ in this equation, we obtain (29). Q.E.D.

It is not possible to develop a policy iteration algorithm for (SM) along the lines of Proposition 4.8 or 4.9. One difficulty is this. If at the kth iteration we have constructed a policy $(\mu_k, \mu_k, \ldots)$, where $\mu_k \in U(C|S)$, then J_{μ_k} is universally measurable but not necessarily lower semianalytic. We would like to find $\mu_{k+1} \in U(C|S)$ such that $T_{\mu_{k+1}}(J_{\mu_k}) \le T(J_{\mu_k}) + \varepsilon$, where $\varepsilon > 0$ is some prescribed small number, but Proposition 7.50 does not apply to this case.

We turn now to the question of convergence of the dynamic programming algorithm under (P). Without additional assumptions, we have only the following result.

Proposition 9.16 (P) We have

$$\bar{J}_\infty \le \bar{J}^*, \tag{32}$$

$$J_\infty \le J^*. \tag{33}$$

Furthermore, the following statements are equivalent:

(a) $\bar{J}_\infty = \bar{T}(\bar{J}_\infty)$,
(b) $\bar{J}_\infty = \bar{J}^*$,
(c) $J_\infty = T(J_\infty)$,
(d) $J_\infty = J^*$.

Proof It is clear that (32) holds and, by Proposition 9.10, implies the equivalence of (a) and (b). Lemma 9.3, Proposition 9.5, and (32) imply (33). Conditions (a) and (c) are equivalent by Lemma 9.3 and Proposition 9.7. Conditions (b) and (d) are equivalent by Lemma 9.3 and Proposition 9.5. Q.E.D.

In Example 1, we have $J_\infty(0) = 0$ and $J^*(0) = \infty$, so strict inequality in (32) and (33) is possible. We present now an example in which not only is J_∞ different from J^*, but J_∞ is Borel-measurable while J^* is not.

EXAMPLE 2 (Blackwell) Let Σ be the set of finite sequences of positive integers and H the set of functions h from Σ into $\{0,1\}$. Then H can be regarded as the countable Cartesian product of copies of $\{0,1\}$ indexed by Σ. Let $\{0,1\}$ have the discrete topology and H the product topology, so H is a complete separable metrizable space (Proposition 7.4). A typical basic open set in H is $\{h \in H | h(s) = 1\ \forall s \in \Sigma_1,\ h(s) = 0\ \forall s \in \Sigma_2\}$, where Σ_1 and Σ_2 are finite subsets of Σ. Consider a Suslin scheme $R: \Sigma \to \mathscr{B}_H$ defined by

$$R(s) = \{h \in H | h(s) = 1\} \qquad \forall s \in \Sigma.$$

Then

$$N(R) = \{h \in H | \exists (\zeta_1, \zeta_2, \ldots) \in \mathscr{N} \text{ such that } h(\zeta_1, \zeta_2, \ldots, \zeta_n) = 1\ \forall n\}$$

is an analytic subset of H (Proposition 7.36). We show with the aid of Appendix B that $N(R)$ is not Borel-measurable. Let Y be an uncountable Borel space and $Q: \Sigma \to \mathscr{B}_Y$ a Suslin scheme such that $N(Q)$ is not Borel-measurable (Proposition B.6). Define $\psi: Y \to H$ by

$$\psi(y)(s) = \begin{cases} 1 & \text{if } y \in Q(s), \\ 0 & \text{if } y \notin Q(s). \end{cases}$$

If Σ_1 and Σ_2 are finite subsets of Σ, then

$$\psi^{-1}(\{h \in H | h(s) = 1\ \forall s \in \Sigma_1,\ h(s) = 0\ \forall s \in \Sigma_2\}) = \left[\bigcap_{s \in \Sigma_1} Q(s)\right] \cap \left[\bigcap_{s \in \Sigma_2} (Y - Q(s))\right]$$

is in $\mathscr{B}_Y$. The collection $\mathscr{E}$ of subsets E of H for which $\psi^{-1}(E) \in \mathscr{B}_Y$ is a σ-algebra containing a base for the topology on H, so, by the remark following Definition 7.6, $\mathscr{E}$ contains $\mathscr{B}_H$ and ψ is Borel-measurable. For each $s \in \Sigma$, we have $Q(s) = \psi^{-1}[R(s)]$, so

$$N(Q) = \bigcup_{z \in \mathscr{N}} \bigcap_{s < z} Q(s) = \bigcup_{z \in \mathscr{N}} \bigcap_{s < z} \psi^{-1}[R(s)] = \psi^{-1}\left[\bigcup_{z \in \mathscr{N}} \bigcap_{s < z} R(s)\right] = \psi^{-1}[N(R)].$$

Since $N(Q)$ is not Borel-measurable, $N(R)$ is also not Borel-measurable.

Define the decision model by taking $S = H\Sigma^*$, where $\Sigma^* = \Sigma \cup \{0\}$, $C = \{1,2,\ldots\}$, $U(x) = C$ for every $x \in S$, and

$$f([h,0],u) = (h,u),$$
$$f([h,(\zeta_1,\zeta_2,\ldots,\zeta_n)],u) = [h,(\zeta_1,\zeta_2,\ldots,\zeta_n,u)].$$

The system transition is deterministic, so the choice of W and $p(dw|x,u)$ is irrelevant. Choose $\alpha = 1$ and

$$g([h,0],u) = \begin{cases} 0 & \text{if } h(u) = 1, \\ 1 & \text{if } h(u) = 0, \end{cases}$$

$$g([h,(\zeta_1,\zeta_2,\ldots,\zeta_n)],u) = \begin{cases} 0 & \text{if } h(\zeta_1,\zeta_2,\ldots,\zeta_n,u) = 1, \\ 1 & \text{if } h(\zeta_1,\zeta_2,\ldots,\zeta_n,u) = 0. \end{cases}$$

If the system begins at $x_0 = [h,0]$ and the horizon is infinite, a positive cost can be avoided if and only if there exists $(\zeta_1,\zeta_2,\ldots)$ such that $h(\zeta_1,\zeta_2,\ldots,\zeta_n) = 1$ for every n, i.e., $J^*([h,0]) = 0$ if and only if $h \in N(R)$. Therefore, J^* is not Borel-measurable. Over the finite horizon, we have

$$J_{k+1}(x) = T(J_k)(x) = \inf_{u \in C}\{g(x,u) + J_k[f(x,u)]\},$$

and since C is countable and f, g, and J_0 are Borel-measurable, J_k is Borel-measurable for $k = 0,1,2,\ldots$. It follows that J_∞ is Borel-measurable.

The equivalent conditions of Proposition 9.16 are not easily verified in practice. We give here some more readily verifiable conditions which imply that $J_\infty = J^*$.

Proposition 9.17 (P)(D) Assume that there exists a nonnegative integer $\bar{k}$ such that for each $x \in S$, $\lambda \in R$, and $k \geq \bar{k}$, the set

$$U_k(x,\lambda) = \left\{ u \in U(x) \middle| g(x,u) + \alpha \int J_k(x')t(dx'|x,u) \leq \lambda \right\} \tag{34}$$

is compact in C. Then $\bar{J}_\infty = \bar{J}^*$, $J_\infty = J^*$, and there exists an optimal nonrandomized stationary policy for (SM).

Proof Under (P), we have, for each k, $J_k \leq J_\infty$, so $J_{k+1} = T(J_k) \leq T(J_\infty)$, and letting $k \to \infty$ we obtain

$$J_\infty \leq T(J_\infty). \tag{35}$$

Let $x \in S$ be such that $J_\infty(x) < \infty$. By Lemma 3.1 for $k \geq \bar{k}$ there exists $u_k \in U(x)$ such that

$$J_{k+1}(x) = g(x,u_k) + \alpha \int J_k(x')t(dx'|x,u_k).$$

Since $J_k \leq J_{k+1} \leq \cdots \leq J_\infty$, it follows that for $k \geq \bar{k}$

$$\begin{aligned} g(x,u_i) + \alpha \int J_k(x')t(dx'|x,u_i) &\leq g(x,u_i) + \alpha \int J_i(x')t(dx'|x,u_i) \\ &= J_{i+1}(x) \leq J_\infty(x) \qquad \forall i \geq k. \end{aligned}$$

Therefore, $\{u_i | i \geq k\} \subset U_k[x, J_\infty(x)]$ for every $k \geq \bar{k}$. Since $U_k[x, J_\infty(x)]$ is

compact, all limit points of the sequence $\{u_i | i \geq k\}$ belong to $U_k[x, J_\infty(x)]$, and at least one such limit point exists. It follows that if $\bar{u}$ is a limit point of the sequence $\{u_i | i \geq \bar{k}\}$, then

$$\bar{u} \in \bigcap_{k=\bar{k}}^{\infty} U_k[x, J_\infty(x)].$$

Therefore, for all $k \geq \bar{k}$,

$$J_\infty(x) \geq g(x, \bar{u}) + \alpha \int J_k(x')t(dx'|x, \bar{u}) \geq J_{k+1}(x).$$

Letting $k \to \infty$ and using the monotone convergence theorem, we obtain

$$J_\infty(x) = g(x, \bar{u}) + \alpha \int J_\infty(x')t(dx'|x, \bar{u}) \geq T(J_\infty)(x) \tag{36}$$

for all $x \in S$ such that $J_\infty(x) < \infty$. We also have that (36) holds if $J_\infty(x) = \infty$, and thus it holds for all $x \in S$. From (35) and (36) we see that $J_\infty = T(J_\infty)$ and conditions (a)–(d) of Proposition 9.16 must hold. In particular, we have from (35) and (36) that for every $x \in S$, there exists $\bar{u} \in U(x)$ such that

$$J^*(x) = g(x, \bar{u}) + \alpha \int J^*(x')t(dx'|x, \bar{u}) = T(J^*)(x).$$

The existence of an optimal nonrandomized stationary policy for (SM) follows from Corollary 9.12.1.

Under (D), conditions (a)–(d) of Proposition 9.16 hold by Proposition 9.14. If we replace g by $g + b$, we obtain a model satisfying (P). This new model also satisfies the hypotheses of the proposition, so there exists an optimal nonrandomized stationary policy for it. This policy is optimal for the original (D) model as well. Q.E.D.

Corollary 9.17.1 (P)(D) Assume that the set $U(x)$ is finite for each $x \in S$. Then $\bar{J}_\infty = \bar{J}^*$, $J_\infty = J^*$, and there exists an optimal nonrandomized stationary policy for (SM). In fact, if C is finite and g and Γ are Borel-measurable, then J^* is Borel-measurable and there exists a Borel-measurable optimal nonrandomized stationary policy for (SM).

Corollary 9.17.2 (P)(D) Suppose conditions (a)–(e) of Definition 8.7 (the lower semicontinuous model) are satisfied. Then $\bar{J}_\infty = \bar{J}^*$, $J_\infty = J^*$, J^* is lower semicontinuous, and there exists a Borel-measurable optimal nonrandomized stationary policy for (SM).

Proof From the proof of Proposition 8.6, we see that J_k is lower semicontinuous for $k = 1, 2, \ldots$, as are the functions

$$\hat{K}_k(x, u) = \begin{cases} g(x, u) + \alpha \int J_k(x')t(dx'|x, u) & \text{if } (x, u) \in \Gamma, \\ \infty & \text{if } (x, u) \notin \Gamma. \end{cases} \tag{37}$$

For $\lambda \in R$ and k fixed, the lower level set

$$\{(x,u) \in SC | \hat{K}_k(x,u) \leq \lambda\} \subset \Gamma$$

is closed, so for each fixed $x \in S$

$$U_k(x,\lambda) = \{u \in C | \hat{K}_k(x,u) \leq \lambda\}$$

is compact. Proposition 9.17 can now be invoked, and it remains only to prove that the optimal nonrandomized stationary policy whose existence is guaranteed by that proposition can be chosen to be Borel-measurable. This will follow from Proposition 9.12 and the proof of Proposition 8.6 once we show that $J_\infty = J^*$ is lower semicontinuous. Under (P), $J_k \uparrow J^*$, so

$$\{x \in S | J^*(x) \leq \lambda\} = \bigcap_{k=0}^{\infty} \{x \in S | J_k(x) \leq \lambda\}$$

is closed, and J^* is lower semicontinuous. Under (D),

$$J_k - b \sum_{j=k}^{\infty} \alpha^k \uparrow J^*,$$

so a similar argument can be used to show that J^* is lower semicontinuous. Q.E.D.

By using the argument used to prove Corollary 8.6.1, we also have the following.

Corollary 9.17.3 The conclusions of Corollary 9.17.2 hold if instead of assuming that C is compact and each Γ^j is closed in Definition 8.7, we assume that each Γ^j is compact.

Proposition 9.17 and its corollaries provide conditions under which the dynamic programming algorithm can be used in the (P) and (D) models to generate J^*. It is also possible to use the dynamic programming algorithm to generate an optimal stationary policy, as is indicated by the next proposition.

Proposition 9.18 (P)(D) Suppose that either $U(x)$ is finite for each $x \in S$ or else conditions (a)–(e) of Definition 8.7 hold. Then for each $k \geq 0$ there exists a universally measurable $\mu_k : S \to C$ such that $\mu_k(x) \in U(x)$ for every $x \in S$ and

$$T_{\mu_k}(J_k) = T(J_k). \tag{38}$$

If $\{\mu_k\}$ is a sequence of such functions, then for each $x \in S$ the sequence $\{\mu_k(x)\}$ has at least one accumulation point. If $\mu : S \to C$ is universally measurable, $\mu(x)$ is an accumulation point of $\{\mu_k(x)\}$ for each $x \in S$ such that $J^*(x) < \infty$, and $\mu(x) \in U(x)$ for each $x \in S$ such that $J^*(x) = \infty$, then $\pi = (\mu, \mu, \ldots)$ is an optimal stationary policy for (SM).

Proof If $U(x)$ is finite for each $x \in S$, then the sets $U_k(x, \lambda)$ of (34) are compact for all $k \geq 0$, $x \in S$, and $\lambda \in R$. The proof of Corollary 9.17.2 shows that these sets are also compact under conditions (a)–(e) of Definition 8.7. The existence of functions $\mu_k : S \to C$ satisfying (38) such that $\mu_k(x) \in U(x)$ for every $x \in S$ is a consequence of Lemma 3.1 and Proposition 7.50.

Under (P) we see from the proof of Proposition 9.17 that $\{\mu_k(x)\}$ has at least one accumulation point for each $x \in S$ such that $J^*(x) < \infty$ and every accumulation point of $\{\mu_k(x)\}$ is in $U(x)$. If $\mu : S \to C$ is universally measurable and $\mu(x)$ is an accumulation point of $\{\mu_k(x)\}$ for each $x \in S$ such that $J^*(x) < \infty$, then from (35), (36), and Proposition 9.17 we have

$$J^*(x) = g[x, \mu(x)] + \alpha \int J^*(x')t(dx'|x, \mu(x)) \tag{39}$$

for all $x \in S$ such that $J^*(x) < \infty$. If $\mu(x) \in U(x)$ for all $x \in S$ such that $J^*(x) = \infty$, then

$$J^*(x) = T(J^*)(x) \leq g[x, \mu(x)] + \alpha \int J^*(x')t(dx'|x, \mu(x)) \leq \infty = J^*(x) \tag{40}$$

for all $x \in S$ such that $J^*(x) = \infty$. From (39) and (40) we have $J^* = T_\mu(J^*)$, and the policy $\pi = (\mu, \mu, \ldots)$ is optimal by Proposition 9.12.

Under (D) we can replace g by $g + b$ to obtain a model satisfying (P) and the hypotheses of the proposition. The conclusions of the proposition are valid for this new model, so they are valid for the original (D) model as well. Q.E.D.

A slightly stronger version of Proposition 9.18 can be found in [S12].

Corollary 9.18.1 If conditions (b)–(e) of Definition 8.7 hold and if each Γ^j of condition (b) is compact, then the conclusions of Proposition 9.18 hold.

9.6 Existence of ε-Optimal Policies

We have characterized stationary optimal policies and given conditions under which optimal policies exist. We turn now to the existence of ε-optimal policies. For fixed $x \in S$, by definition there is a policy which is ε-optimal at x. We would like to know how this collection of policies, each of which is ε-optimal at a single point, can be pieced together to form a single policy which is ε-optimal at every point. There is a related question concerning optimal policies. If at each point there is a policy which is optimal at that point, is it possible to find an optimal policy? Answers to these questions are provided by the next two propositions.

Proposition 9.19 (P)(D) For each $\varepsilon > 0$, there exists an ε-optimal nonrandomized Markov policy for (SM), and if $\alpha < 1$, it can be taken to be

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

Proof Choose $\varepsilon > 0$ and $\varepsilon_k > 0$ such that $\sum_{k=0}^{\infty} \alpha^k \varepsilon_k = \varepsilon$. If $\alpha < 1$, let $\varepsilon_k = (1 - \alpha)\varepsilon$ for every k. By Proposition 7.50, there are universally measurable functions $\mu_k : S \to C$, $k = 0, 1, \ldots$, such that $\mu_k(x) \in U(x)$ for every $x \in S$ and

$$T_{\mu_k}(J^*) \leq J^* + \varepsilon_k.$$

If $\alpha < 1$, we choose all the μ_k to be identical. Then

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

Continuing this process, we have

$$(T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J^*) \leq J^* + \sum_{j=0}^{k} \alpha^j \varepsilon_j \leq J^* + \varepsilon,$$

and, letting $k \to \infty$, we obtain

$$\lim_{k \to \infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J^*) \leq J^* + \varepsilon.$$

Under (P) we have

$$J_\pi = \lim_{k \to \infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J_0) \leq \lim_{k \to \infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J^*), \tag{41}$$

so $\pi = (\mu_0, \mu_1, \ldots)$ is ε-optimal. Under (D),

$$\begin{aligned} J_0 &\leq J^* + [b/(1 - \alpha)], \\ (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J_0) &\leq (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J^* + [b/(1 - \alpha)]) \\ &= [\alpha^{k+1} b/(1 - \alpha)] + (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})(J^*), \end{aligned}$$

so (41) is valid and $\pi = (\mu_0, \mu_1, \ldots)$ is ε-optimal. This proves the first part of the proposition.

Suppose that for each $x \in S$ there is a policy for (SM) which is optimal at x. Fix x and let $\pi = (\mu_0, \mu_1, \ldots)$ be a policy which is optimal at x. By Proposition 9.1, we may assume without loss of generality that π is Markov. By Lemma 8.4(b) and (c), we have

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

Consequently,

$$T_{\mu_0}(J^*)(x) = T(J^*)(x).$$

This implies that the infimum in the expression

$$\inf_{u \in U(x)} \left\{ g(x,u) + \alpha \int J^*(x') t(dx'|x,u) \right\}$$

is achieved. Since x is arbitrary, Corollary 9.12.1 implies the existence of an optimal nonrandomized stationary policy. Q.E.D.

Proposition 9.20 (N) For each $\varepsilon > 0$, there exists an ε-optimal nonrandomized semi-Markov policy for (SM). If for each $x \in S$ there exists a policy for (SM) which is optimal at x, then there exists a semi-Markov (randomized) optimal policy.

Proof Under (N) we have $J_k \downarrow J^*$ (Proposition 9.14), so, given $\varepsilon > 0$, the analytically measurable sets

$$\begin{aligned} A_k = &\{x \in S | J^*(x) > -\infty, J_k(x) \le J^*(x) + \varepsilon/2\} \\ &\cup \{x \in S | J^*(x) = -\infty, J_k(x) \le -(2+\varepsilon^2)/2\varepsilon\} \end{aligned}$$

converge up to S as $k \to \infty$. By Proposition 8.3, for each k there exists a k-stage nonrandomized semi-Markov policy π^k such that for every $x \in S$

$$J_{k,\pi^k}(x) \le \begin{cases} J_k(x) + (\varepsilon/2) & \text{if } J_k(x) > -\infty, \\ -1/\varepsilon & \text{if } J_k(x) = -\infty. \end{cases}$$

Then for $x \in A_k$ we have either $J^*(x) > -\infty$ and

$$J_{k,\pi^k}(x) \le J_k(x) + (\varepsilon/2) \le J^*(x) + \varepsilon,$$

or else $J^*(x) = -\infty$. If $J^*(x) = -\infty$, then either $J_k(x) = -\infty$ and

$$J_{k,\pi^k}(x) \le -1/\varepsilon,$$

or else $J_k(x) > -\infty$ and

$$J_{k,\pi^k}(x) \le J_k(x) + (\varepsilon/2) \le -[(2+\varepsilon^2)/2\varepsilon] + (\varepsilon/2) = -1/\varepsilon.$$

Choose any $\mu \in U(C|S)$ and define $\hat{\pi}^k = (\mu_0^k, \ldots, \mu_{k-1}^k, \mu, \mu, \ldots)$, where $\pi^k = (\mu_0^k, \ldots, \mu_{k-1}^k)$. For every $x \in A_k$, we have

$$J_{\hat{\pi}^k}(x) \le J_{k,\pi^k}(x) \le \begin{cases} J^*(x) + \varepsilon & \text{if } J^*(x) > -\infty, \\ -1/\varepsilon & \text{if } J^*(x) = -\infty, \end{cases}$$

so $\hat{\pi}^k$ is a nonrandomized semi-Markov policy which is ε-optimal for every $x \in A_k$. The policy π defined to be $\hat{\pi}^k$ when the initial state is in A_k, but not in A_j for any $j < k$, is semi-Markov, nonrandomized, and ε-optimal at every $x \in \bigcup_{k=1}^{\infty} A_k = S$.

Suppose now that for each $x \in S$ there exists a policy π^x for (SM) which is optimal at x. Let $\bar{\pi}^x$ be a policy for (DM) which corresponds to π^x, and let $(p_x, q_0^x, q_1^x, \ldots)$ be the sequence generated from p_x by $\bar{\pi}^x$ via (10) and (11). If $G: \Delta \to [-\infty, 0]$ is defined by (15), then we have from Proposition 9.3 that

$$J^*(x) = J_{\pi^x}(x) = \bar{J}_{\bar{\pi}^x}(p_x) = G(p_x, q_0^x, q_1^x, \ldots). \tag{42}$$

We have from Proposition 9.5 and (16) that

$$J^*(x) = \bar{J}^*(p_x) = \inf_{(q_0, q_1, \ldots) \in \Delta_{p_x}} G(p_x, q_0, q_1, \ldots). \tag{43}$$

Therefore the infimum in (43) is attained for every $p_x \in \bar{S}$, where $\bar{S} = \{p_y | y \in S\}$, so by Proposition 7.50, there exists a universally measurable selector $\psi: \bar{S} \to P(SC)P(SC)\cdots$ such that $\psi(p_x) \in \Delta_{p_x}$ and

$$J^*(x) = \bar{J}^*(p_x) = G[p_x, \psi(p_x)] \qquad \forall p_x \in \bar{S}.$$

Let $\delta: S \to \bar{S}$ be the homeomorphism $\delta(x) = p_x$ and let $\varphi(x) = \psi[\delta(x)]$. Then φ is universally measurable, $\varphi(x) \in \Delta_{p_x}$, and

$$J^*(x) = G[p_x, \varphi(x)] \qquad \forall x \in S. \tag{44}$$

Denote

$$\varphi(x) = [q_0(d(x_0, u_0)|x),\ q_1(d(x_1, u_1)|x), \ldots].$$

For each $k \geq 0$, $q_k(d(x_k, u_k)|x)$ is a universally measurable stochastic kernel on $S_k C_k$ given S, and by Proposition 7.27 and Lemma 7.28(a), (b), $q_k(d(x_k, u_k)|x)$ can be decomposed into its marginal $p_k(dx_k|x)$, which is a universally measurable stochastic kernel on S_k given S, and a universally measurable stochastic kernel $\mu_k(du_k|x, x_k)$ on C_k given SS_k. Since $p_0(dx_0|x) = p_x(dx_0)$, the stochastic kernel $\mu_0(du_0|x, x_0)$ is arbitrary except when $x = x_0$. Set

$$\bar{\mu}_0(du_0|x) = \mu_0(du_0|x, x) \qquad \forall x \in S.$$

The sequence $\pi = (\bar{\mu}_0, \mu_1, \mu_2, \ldots)$ is a randomized semi-Markov policy for (SM). From (7) of Chapter 8, we see that for each $x \in S$

$$q_k(\pi, p_x) = q_k(d(x_k, u_k)|x) \qquad \forall x \in S, \quad k = 0, 1, \ldots.$$

From (5), (15), and (44), we have

$$J_\pi(x) = G[p_x, q_0(\pi, p_x), q_1(\pi, p_x), \ldots] = J^*(x) \qquad \forall x \in S,$$

so π is optimal. Q.E.D.

Although randomized polcies may be considered inferior and are avoided in practice, under (N) as posed here they cannot be disregarded even in deterministic problems, as the following example demonstrates.

EXAMPLE 3 (St. Petersburg paradox) Let $S = \{0, 1, 2, \ldots\}$, $C = \{0, 1\}$, $U(x) = C$ for every $x \in S$, $\alpha = 1$,

$$f(x, u) = \begin{cases} x + 1 & \text{if } u = 1, \quad x \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$g(x, u) = \begin{cases} -2^x & \text{if } x \neq 0, \quad u = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Beginning in state one, any nonrandomized policy either increases the state by one indefinitely and incurs no nonzero cost or else, after k increases, jumps the system to zero at a cost of -2^{k+1}, where it remains at no further cost. Thus $J^*(1) = -\infty$, but this cost is not achieved by any nonrandomized policy. On the other hand, the randomized stationary policy which jumps the system to zero with probability $\frac{1}{2}$ when the state x is nonzero yields an expected cost of $-\infty$ and is optimal at every $x \in S$.

The one-stage cost g in Example 3 is unbounded, but by a slight modification an example can be constructed in which g is bounded and the only optimal policies are randomized. If one stipulates that J^* must be finite, it may be possible to restrict attention to nonrandomized policies in Proposition 9.20. This is an unsolved problem.

If (SM) is lower semicontinuous, then Proposition 9.19 can be strengthened, as Corollary 9.17.2 shows. Similarly, if (SM) is upper semicontinuous, a stronger version of Proposition 9.20 can be proved.

Proportional 9.21 Assume (SM) satisfies conditions (a)–(d) of Definition 8.8 (the upper semicontinuous model).

(D) For each $\varepsilon > 0$, there exists a Borel-measurable, ε-optimal, nonrandomized, stationary policy.

(N) For each $\varepsilon > 0$, there exists a Borel-measurable, ε-optimal, nonrandomized, semi-Markov policy.

Under both (D) and (N), J^* is upper semicontinuous.

Proof Under (D) and (N) we have $\lim_{k \to \infty} J_k = J^*$ (Proposition 9.14), and each J_k is upper semicontinuous (Proposition 8.7). By an argument similar to that used in the proof of Corollary 9.17.2, J^* is upper semicontinuous.

By using Proposition 7.34 in place of Proposition 7.50, the proof of Proposition 9.19 can be modified to show the existence of a Borel-measurable, ε-optimal, nonrandomized, stationary policy under (D). By using Proposition 8.7 in place of Proposition 8.3, the proof of Proposition 9.20 can be modified to show the existence of a Borel-measurable, ε-optimal, nonrandomized, semi-Markov policy under (N). Q.E.D.

Chapter 10

The Imperfect State Information Model

In the models of Chapters 8 and 9 the current state of the system is known to the controller at each stage. In many problems of practical interest, however, the controller has instead access only to imperfect measurements of the system state. This chapter is devoted to the study of models relating to such situations. In our analysis we will encounter nonstationary versions of the models of Chapters 8 and 9. We will show in the next section that nonstationary models can be reduced to stationary ones by appropriate reformulation. We will thus be able to obtain nonstationary counterparts to the results of Chapters 8 and 9.

10.1 Reduction of the Nonstationary Model—State Augmentation

The finite horizon stochastic optimal control model of Definition 8.1 and the infinite horizon stochastic optimal control model of Definition 9.1 are said to be *stationary*, i.e., the data defining the model does not vary from stage to stage. In this section we define a nonstationary model and show how it can be reduced to a stationary one by augmenting the state with the time index.

We combine the treatments of the finite and infinite horizon models. Thus when $N = \infty$ and notation of the form $S_0, S_1, \ldots, S_{N-1}$ or $k = 0, \ldots, N-1$ appears, we take this to mean $S_0, S_1, \ldots$ and $k = 0, 1, \ldots$, respectively.

Definition 10.1 A *nonstationary stochastic optimal control model*, denoted by (NSM), consists of the following objects:

N *Horizon.* A positive integer or ∞.

S_k, $k = 0, \ldots, N-1$ *State spaces.* For each k, S_k is a nonempty Borel space.

C_k, $k = 0, \ldots, N-1$ *Control spaces.* For each k, C_k is a nonempty Borel space.

U_k, $k = 0, \ldots, N-1$ *Control constraints.* For each k, U_k is a function from S_k to the set of nonempty subsets of C_k, and the set

$$\Gamma_k = \{(x_k, u_k) | x_k \in S_k,\ u_k \in U_k(x_k)\} \tag{1}$$

is analytic in $S_k C_k$.

W_k, $k = 0, \ldots, N-1$ *Disturbance spaces.* For each k, W_k is a nonempty Borel space.

$p_k(dw_k | x_k, u_k)$, $k = 0, \ldots, N-1$ *Disturbance kernels.* For each k, $p_k(dw_k | x_k, u_k)$ is a Borel-measurable stochastic kernel on W_k given $S_k C_k$.

f_k, $k = 0, \ldots, N-2$ *System functions.* For each k, f_k is a Borel-measurable function from $S_k C_k W_k$ to S_{k+1}.

α *Discount factor.* A positive real number.

g_k, $k = 0, \ldots, N-1$ *One-stage cost functions.* For each k, g_k is a lower semianalytic function from Γ_k to R^*.

We envision a system which begins at some $x_k \in S_k$ and moves successively through state spaces $S_{k+1}, S_{k+2}, \ldots$ and, if $N < \infty$, finally terminates in S_{N-1}. A policy governing such a system evolution is a sequence $\pi^k = (\mu_k, \mu_{k+1}, \ldots, \mu_{N-1})$, where each μ_j is a universally measurable stochastic kernel on C_j given $S_k C_k \cdots C_{j-1} S_j$ satisfying

$$\mu_j(U_j(x_j) | x_k, u_k, \ldots, u_{j-1}, x_j) = 1$$

for every $(x_k, u_k, \ldots, u_{j-1}, x_j)$. Such a policy is called a *k-originating policy* and the collection of all k-originating policies will be denoted by Π^k. The concepts of *semi-Markov*, *Markov*, *nonrandomized* and *$\mathscr{F}$-measurable* policies are analogous to those of Definitions 8.2 and 9.2. The set Π^0 is also written as Π', and the subset of Π' consisting of all Markov policies is denoted by Π.

Define the Borel-measurable *state transition stochastic kernels* by

$$t_k(\underline{S}_{k+1} | x_k, u_k) = p_k(\{w_k \in W_k | f_k(x_k, u_k, w_k) \in \underline{S}_{k+1}\} | x_k, u_k) \qquad \forall \underline{S}_{k+1} \in \mathscr{B}_{S_{k+1}}.$$

Given a probability measure $p_k \in P(S_k)$ and a policy $\pi^k = (\mu_k, \ldots, \mu_{N-1}) \in \Pi^k$, define for $j = k, k+1, \ldots, N-1$

$$q_j(\pi^k, p_k)(\underline{S}_j \underline{C}_j) = \int_{S_k} \int_{C_k} \cdots \int_{C_{j-1}} \int_{\underline{S}_j} \mu_j(\underline{C}_j | x_k, u_k, \ldots, u_{j-1}, x_j) \times t_{j-1}(dx_j | x_{j-1}, u_{j-1}) \times \mu_{j-1}(du_{j-1} | x_k, u_k, \ldots, u_{j-2}, x_{j-1}) \cdots \mu_k(du_k | x_k) p_k(dx_k) \quad \forall \underline{S}_j \in \mathcal{B}_{S_j}, \quad \underline{C}_j \in \mathcal{B}_{C_j}. \tag{2}$$

There is a unique probability measure $q_j(\pi^k, p_k) \in P(S_j C_j)$ satisfying (2).

If the horizon N is finite, we treat (NSM) only under one of the following assumptions:

$$\int_{S_j C_j} g_j^-(x_j, u_j)\, dq_j(\pi^k, p_{x_k}) < \infty \qquad \forall \pi^k \in \Pi^k, \quad x_k \in S_k, \quad k \le j \le N-1, \quad k = 0, \ldots, N-1. \tag{F$^+$}$$

$$\int_{S_j C_j} g_j^+(x_j, u_j)\, dq_j(\pi^k, p_{x_k}) < \infty \qquad \forall \pi^k \in \Pi^k, \quad x_k \in S_k, \quad k \le j \le N-1, \quad k = 0, \ldots, N-1. \tag{F$^-$}$$

If $N = \infty$, we treat (NSM) only under one of the assumptions:

(P) $0 \le g_k(x_k, u_k)$ for every $(x_k, u_k) \in \Gamma_k$, $k = 0, \ldots, N-1$.

(N) $g_k(x_k, u_k) \le 0$ for every $(x_k, u_k) \in \Gamma_k$, $k = 0, \ldots, N-1$.

(D) $0 < \alpha < 1$, and for some $b \in R$, $-b \le g_k(x_k, u_k) \le b$ for every $(x_k, u_k) \in \Gamma_k$, $k = 0, \ldots, N-1$.

As in Chapters 8 and 9, the symbols (F$^+$), (F$^-$), (P), (N), and (D) will be used to indicate when a result is valid under the appropriate assumption.

We define the *k-originating cost corresponding to* π^k at $x_k \in S_k$ to be

$$J_{\pi^k}(x_k, k) = \sum_{j=k}^{N-1} \alpha^j \int_{S_j C_j} g_j(x_j, u_j)\, dq_j(\pi^k, p_{x_k}),$$

and the *k-originating optimal cost* at $x_k \in S_k$ to be

$$J^*(x_k, k) = \inf_{\pi^k \in \Pi^k} J_{\pi^k}(x_k, k). \tag{3}$$

A policy $\pi \in \Pi^0$ is *ε-optimal* at $x_0 \in S_0$ if

$$J_\pi(x_0, 0) \le \begin{cases} J^*(x_0, 0) + \varepsilon & \text{if} \quad J^*(x_0, 0) > -\infty, \\ -1/\varepsilon & \text{if} \quad J^*(x_0, 0) = -\infty. \end{cases}$$

The policy π is *optimal* at x_0 if $J_\pi(x_0, 0) = J^*(x_0, 0)$. We say $\pi \in \Pi^0$ is *ε-optimal* (*optimal*) if it is ε-optimal (optimal) at every $x_0 \in S_0$. Let $\{\varepsilon_n\}$ be a sequence of positive numbers with $\varepsilon_n \downarrow 0$. A sequence of policies $\{\pi_n\} \subset \Pi^0$ is said to

exhibit $\{\varepsilon_n\}$-*dominated convergence to optimality* if

$$\lim_{n\to\infty} J_{\pi_n}(x_0,0) = J^*(x_0,0) \qquad \forall x_0 \in S_0,$$

and for $n = 2,3,\ldots$

$$J_{\pi_n}(x_0,0) \leq \begin{cases} J^*(x_0,0) + \varepsilon_n & \text{if } J^*(x_0,0) > -\infty, \\ J_{\pi_{n-1}}(x_0,0) + \varepsilon_n & \text{if } J^*(x_0,0) = -\infty. \end{cases}$$

Definition 10.2 Let a nonstationary stochastic optimal control model as defined by Definition 10.1 be given. The corresponding *stationary stochastic optimal control model*, denoted by (SSM), consists of the following objects. (T is both a terminal state and the only control available at that state. If $N = \infty$, the introduction of T is unnecessary.):

$S = \bigcup_{k=0}^{N-1} \{(x_k,k) | x_k \in S_k\} \cup \{T\}$ *State space.*

$C = \bigcup_{k=0}^{N-1} \{(u_k,k) | u_k \in C_k\} \cup \{T\}$ *Control space.*

U *Control constraint.* A function from S to the set of nonempty subsets of C defined by $U(x_k,k) = \{(u_k,k) | u_k \in U_k(x_k)\}$, $U(T) = \{T\}$.

$W = \bigcup_{k=0}^{N-1} \{(w_k,k) | w_k \in W_k\}$ *Disturbance space.*

$p(dw|x,u)$ *Disturbance kernel.* If $\underline{W}_k \in \mathcal{B}_{W_k}$, we define

$$p[\{(w_k,k) | w_k \in \underline{W}_k\} | (x_k,k),(u_k,k)] = p_k(\underline{W}_k | x_k,u_k). \tag{4}$$

f *System function.* We define for $k = 0,\ldots,N-2$

$$f[(x_k,k),(u_k,k),(w_k,k)] = [f_k(x_k,u_k,w_k), k+1], \tag{5}$$

and for the remaining two stages

$$f[(x_{N-1},N-1),(u_{N-1},N-1),(w_{N-1},N-1)] = T, \tag{6}$$

$$f(T,T,w) = T. \tag{7}$$

α *Discount factor.*

g *One-stage cost function.* We define

$$g[(x_k,k),(u_k,k)] = g_k(x_k,u_k), \tag{8}$$

$$g(T,T) = 0. \tag{9}$$

N *Horizon.*

Consider the mapping $\varphi_k : S_k \to S$ given by $\varphi_k(x_k) = (x_k,k)$. We endow S with the topology that makes each φ_k a homeomorphism, and we endow C and W with similar topologies. The spaces S, C, and W are Borel. The set

$$\begin{aligned} \Gamma &= \{(x,u) | x \in S, u \in U(x)\} \\ &= \bigcup_{k=0}^{N-1} \{[(x_k,k),(u_k,k)] | (x_k,u_k) \in \Gamma_k\} \cup \{(T,T)\} \end{aligned}$$

is analytic, and g defined on Γ by (8) and (9) is lower semianalytic. The disturbance kernel $p(dw|x,u)$ is not defined on all of SC by (4), but it is defined on a Borel subset of SC containing $\Gamma - \{(T,T)\}$, which is all that is necessary. Likewise, the system function f is not defined on all of SCW by (5)–(7), but the set of points where it is not defined has probability zero under any policy governing the system evolution. Both $p(dw|x,u)$ and f are Borel-measurable on their domains. Thus (SSM) is a special case of the stochastic optimal control model of Definition 8.1 ($N < \infty$) or Definition 9.1 ($N = \infty$).

If $N < \infty$, the (F^+) and (F^-) assumptions on (SSM) are given in Section 8.1. These are equivalent to the respective (F^+) and (F^-) assumptions on (NSM) given earlier in this section. If $N = \infty$, the (P), (N), and (D) assumptions on (SSM) of Definition 9.1 are equivalent to the respective (P), (N), and (D) assumptions on (NSM) given earlier in this section.

The reader can verify that there is a correspondence of policies between (NSM) and (SSM), and the optimal cost at $(x_k,k)\in S$ for (SSM) is $J^*(x_k,k)$ given by (3). Because of these facts, results already proved for (SSM) with either a finite or infinite horizon have immediate counterparts for (NSM). An illustration of this is the nonstationary optimality equation.

Proposition 10.1 (P)(N)(D) Let $J^*(x_k,k)$ be defined by (3). For fixed k, $J^*(x_k,k)$ is lower semianalytic on S_k, and

$$J^*(x_k,k) = \inf_{u_k \in U_k(x_k)} \left\{ g_k(x_k,u_k) + \alpha \int_{S_{k+1}} J^*(x_{k+1},k+1) t_k(dx_{k+1}|x_k,u_k) \right\}.$$

We do not list all the results for (NSM) that can be obtained from (SSM). The reader may verify, for example, that the existence results of Propositions 8.3 and 8.4, are valid for (NSM) in exactly the form stated. From Propositions 9.19 and 9.20 we conclude that, under (P) and (D), an ε-optimal nonrandomized Markov policy exists for (NSM), while under (N), an ε-optimal nonrandomized semi-Markov policy exists. In what follows, we make use of these results and reference only their stationary versions.

10.2 Reduction of the Imperfect State Information Model—Sufficient Statistics

Before defining the imperfect state information model, we give without proof some of the standard properties of conditional expectations and probabilities we will be using. For a detailed treatment, see Ash [A1]. Throughout this discussion, $(\Omega, \mathcal{F}, P)$ is a probability space and X is an extended real-valued random variable on Ω for which either $E[X^+]$ or $E[X^-]$ is finite.

If $\mathcal{D} \subset \mathcal{F}$ is a σ-algebra on Ω, then the *expectation of X conditioned on $\mathcal{D}$* is any $\mathcal{D}$-measurable, extended real-valued, random variable $E[X|\mathcal{D}](\cdot)$

on Ω which satisfies

$$\int_D X(\omega)P(d\omega) = \int_D E[X|\mathscr{D}](\omega)P(d\omega) \qquad \forall D \in \mathscr{D}.$$

It can be shown that at least one such random variable exists. Any such random variable will be called a *version* of $E[X|\mathscr{D}]$. If $X(\omega) \geq b$ for some $b \in R$ and every $\omega \in \Omega$, then it can be shown that for any version $E[X|\mathscr{D}](\cdot)$ the random variable $\hat{E}[X|\mathscr{D}](\cdot)$ defined by

$$\hat{E}[X|\mathscr{D}](\omega) = \max\{E[X|\mathscr{D}](\omega), b\},$$

is also a version of $E[X|\mathscr{D}]$. If $\mathscr{E} \subset \mathscr{D}$ is a collection of sets which is closed under finite intersections and generates the σ-algebra $\mathscr{D}$ and if Y is an extended real-valued, $\mathscr{D}$-measurable, random variable satisfying

$$\int_D X(\omega)P(d\omega) = \int_D Y(\omega)P(d\omega) \qquad \forall D \in \mathscr{E}, \tag{10}$$

then Y satisfies (10) for every $D \in \mathscr{D}$, and Y is a version of $E[X|\mathscr{D}]$. If $\mathscr{E} \subset \mathscr{D}$ is a σ-algebra, then

$$E\{E[X|\mathscr{D}]|\mathscr{E}\}(\omega) = E[X|\mathscr{E}](\omega) \tag{11}$$

for P almost every ω.

Suppose now that $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ are measurable spaces and $Y_1 : \Omega \to \Omega_1$ and $Y_2 : \Omega \to \Omega_2$ are measurable. Let $g : \Omega_1\Omega_2 \to R^*$ be measurable and satisfy either $E[g^+(Y_1, Y_2)] < \infty$ or $E[g^-(Y_1, Y_2)] < \infty$. We define

$$E[X|Y_1](\omega) = E[X|\mathscr{F}(Y_1)](\omega),$$

where

$$\mathscr{F}(Y_1) = \{Y_1^{-1}(F)|F \in \mathscr{F}_1\}.$$

We define for $y_1 \in \Omega_1$

$$E[X|Y_1 = y_1] = E[X|Y_1](\omega(y_1)),$$

where $\omega(y_1)$ is any element of $Y_1^{-1}(\{y_1\})$. Since $E[X|Y_1]$ is $\mathscr{F}(Y_1)$-measurable, it is constant on $Y_1^{-1}(\{y_1\})$, and this definition makes sense. Note that $E[X|Y_1 = y_1]$ is a function of y_1, not of ω. We have for any $y_1 \in Y_1$

$$E[g(Y_1, Y_2)|Y_1 = y_1] = E[g(y_1, Y_2)] \tag{12}$$

for P almost every y_1. We use the phrase "for P almost every y_1" to indicate that, in this case,

$$P(\{\omega \in \Omega | (12) \text{ fails when } y_1 = Y_1(\omega)\}) = 0.$$

For $F \in \mathscr{F}_2$, define

$$P[Y_2 \in F|Y_1](\omega) = E[\chi_F(Y_2)|Y_1](\omega),$$
$$P[Y_2 \in F|Y_1 = y_1] = E[\chi_F(Y_2)|Y_1 = y_1].$$

Suppose $t(dy_2|y_1)$ is a stochastic kernel on $(\Omega_2, \mathscr{F}_2)$ given Ω_1 such that for every $F \in \mathscr{F}_2$

$$P[Y_2 \in F | Y_1 = y_1] = t(F|y_1)$$

for P almost every y_1. Then (12) can be extended to

$$E[g(Y_1, Y_2)|Y_1 = y_1] = E[g(y_1, Y_2)] = \int g(y_1, y_2)t(dy_2|y_1) \tag{13}$$

for P almost every y_1. We will find (11) and (13) particularly useful in our treatment of the imperfect state information model. They will be used without reference to this discussion.

Definition 10.3 The *imperfect state information stochastic optimal control model* (ISI) is the ten-tuple $(S, C, (U_0, \ldots, U_{N-1}), Z, \alpha, g, t, s_0, s, N)$ described as follows:

S, C, α, g, t *State space, control space, discount factor, one-stage cost function,* and *state transition kernel* as given in Definition 8.1 and (3) of Chapter 8. We assume that g is defined on all of SC.

Z *Observation space.* A nonempty Borel space.

$U_k, k = 0, \ldots, N-1$ *Control constraints.* Define for $k = 0, \ldots, N-1$,

$$I_k = Z_0 C_0 \cdots C_{k-1} Z_k. \tag{14}$$

An element of I_k is called a *kth information vector.* For each k, U_k is a mapping from I_k to the set of nonempty subsets of C such that

$$\Gamma_k = \{(i_k, u) | i_k \in I_k, u \in U_k(i_k)\} \tag{15}$$

is analytic.

s_0 *Initial observation kernel.* A Borel-measurable stochastic kernel on Z given S.

s *Observation kernel.* A Borel-measurable stochastic kernel on Z given CS.

N *Horizon.* A positive integer or ∞.

For the sake of simplicity, we have eliminated the system function, disturbance space, and disturbance kernel from the model definition. In what follows, our notation will generally indicate a finite N. If $N = \infty$, the appropriate interpretation is required.

The system moves stochastically from state x_k to state x_{k+1} via the state transition kernel $t(dx_{k+1}|x_k, u_k)$ and generates cost at each stage of $g(x_k, u_k)$. The observation z_{k+1} is stochastically generated via the observation kernel $s(dz_{k+1}|u_k, x_{k+1})$ and added to the past observations and controls $(z_0, u_0, \ldots, z_k, u_k)$ to form the $(k+1)$st information vector $i_{k+1} = (z_0, u_0, \ldots, z_k, u_k, z_{k+1})$. The first information vector $i_0 = (z_0)$ is generated by the initial observation

kernel $s_0(dz_0|x_0)$, and the initial state x_0 has some given initial distribution p. The goal is to choose u_k dependent on the kth information vector i_k so as to minimize

$$E\left\{\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k)\right\}.$$

Definition 10.4 A *policy* for (ISI) is a sequence $\pi = (\mu_0, \ldots, \mu_{N-1})$ such that, for each k, $\mu_k(du_k|p; i_k)$ is a universally measurable stochastic kernel on C given $P(S)I_k$ satisfying

$$\mu_k(U_k(i_k)|p; i_k) = 1 \qquad \forall (p; i_k) \in P(S)I_k.$$

If for each p, k, and i_k, $\mu_k(du_k|p; i_k)$ assigns mass one to some point in C, π is *nonrandomized.*

The concepts of Markov and semi-Markov policies are of no use in (ISI), since the initial distribution, past observations, and past controls are of genuine value in estimating the current state. Thus we expect policies to depend on the initial distribution p and the total information vector. In the remainder of this chapter, Π will denote the set of all policies in (ISI).

Just as we denote the set of all sequences of the form $(z_0, u_0, \ldots, u_{k-1}, z_k) \in ZC \cdots CZ$ by I_k and call these sequences the kth information vectors, we find it notationally convenient to denote the set of all sequences of the form $(x_0, z_0, u_0, \ldots, x_k, z_k, u_k) \in SZC \cdots SZC$ by H_k and call these sequences the kth *history vectors.* Except for u_k, the kth information vector is that portion of the kth history vector known to the controller at the kth stage. Given $p \in P(S)$ and $\pi = (\mu_0, \ldots, \mu_{N-1}) \in \Pi$, by Proposition 7.45 there is a sequence of consistent probability measures $P_k(\pi, p)$ on H_k, $k = 0, \ldots, N-1$, defined on measurable rectangles by

$$\begin{aligned} &P_k(\pi, p)(\underline{S}_0 \underline{Z}_0 \underline{C}_0 \cdots \underline{S}_k \underline{Z}_k \underline{C}_k) \\ &\quad = \int_{\underline{S}_0} \int_{\underline{Z}_0} \int_{\underline{C}_0} \cdots \int_{\underline{S}_k} \int_{\underline{Z}_k} \mu_k(\underline{C}_k|p; z_0, u_0, \ldots, u_{k-1}, z_k) s(dz_k|u_{k-1}, x_k) \\ &\qquad \times t(dx_k|u_{k-1}, x_{k-1}) \cdots \mu_0(du_0|p; z_0) s_0(dz_0|x_0) p(dx_0). \end{aligned} \tag{16}$$

Definition 10.5 Given $p \in P(S)$, a policy $\pi = (\mu_0, \ldots, \mu_{N-1}) \in \Pi$, and a positive integer $K \leq N$, the *K-stage cost corresponding to π* at p is

$$J_{K,\pi}(p) = \int_{H_{K-1}} \left[\sum_{k=0}^{K-1} \alpha^k g(x_k, u_k)\right] dP_{K-1}(\pi, p). \tag{17}$$

If $N < \infty$, the *cost corresponding to π* is $J_{N,\pi}$, and we assume either

$$\int_{H_{N-1}} \left[\sum_{k=0}^{N-1} \alpha^k g^-(x_k, u_k)\right] dP_{N-1}(\pi, p) < \infty \qquad \forall \pi \in \Pi, \quad p \in P(S) \quad (\mathrm{F}^+)$$

or

$$\int_{H_{N-1}} \left[\sum_{k=0}^{N-1} \alpha^k g^+(x_k, u_k) \right] dP_{N-1}(\pi, p) < \infty \qquad \forall \pi \in \Pi, \quad p \in P(S). \quad (\mathrm{F}^-)$$

If $N = \infty$, the *cost corresponding to* π is $J_\pi = \lim_{K \to \infty} J_{K,\pi}$, and to ensure that this limit is a well-defined extended real number, we impose one of the following conditions:

(P) $0 \leq g(x, u)$ for every $(x, u) \in SC$.

(N) $g(x, u) \leq 0$ for every $(x, u) \in SC$.

(D) $0 < \alpha < 1$, and for some $b \in R$, $-b \leq g(x, u) \leq b$ for every $(x, u) \in SC$.

The *optimal cost* at p is

$$J_N^*(p) = \inf_{\pi \in \Pi} J_{N,\pi}(p).$$

The concepts of *optimality at p*, *optimality*, *ε-optimality at p*, and *ε-optimality* of policies are analogous to those given in Definition 8.3.

If $N < \infty$ and (F^+) or (F^-) holds, then by Lemma 7.11(b)

$$J_{N,\pi}(p) = \sum_{k=0}^{N-1} \alpha^k \int_{H_k} g(x_k, u_k)\, dP_k(\pi, p) \qquad \forall \pi \in \Pi, \quad p \in P(S). \qquad (18)$$

If $N = \infty$ and (P), (N), or (D) holds, then

$$J_\pi(p) = \sum_{k=0}^{\infty} \alpha^k \int_{H_k} g(x_k, u_k)\, dP_k(\pi, p) \qquad \forall \pi \in \Pi, \quad p \in P(S). \qquad (19)$$

To aid in the analysis of (ISI), we introduce the idea of a statistic sufficient for control. This statistic is defined in such a way that knowledge of its values is sufficient to control the model.

Definition 10.6 A *statistic* for the model (ISI) is a sequence $(\eta_0, \ldots, \eta_{N-1})$ of Borel-measurable functions $\eta_k : P(S) I_k \to Y_k$, where Y_k is a nonempty Borel space, $k = 0, \ldots, N-1$. The statistic $(\eta_0, \ldots, \eta_{N-1})$ is *sufficient for control* provided:

(a) For each k, there exists an analytic set $\hat{\Gamma}_k \subset Y_k C$ such that $\mathrm{proj}_{Y_k}(\hat{\Gamma}_k) = Y_k$ and for every $p \in P(S)$

$$\Gamma_k = \{(i_k, u) \mid [\eta_k(p; i_k), u] \in \hat{\Gamma}_k\}, \qquad (20)$$

where Γ_k is defined by (15). We define

$$\hat{U}_k(y_k) = (\hat{\Gamma}_k)_{y_k}. \qquad (21)$$

(b) There exist Borel-measurable stochastic kernels $\hat{t}_k(dy_{k+1} | y_k, u_k)$ on Y_{k+1} given $Y_k C$ such that for every $p \in P(S)$, $\pi \in \Pi$, $\underline{Y}_{k+1} \in \mathscr{B}_{Y_{k+1}}$, $k = 0, \ldots,$

$N-2$, we have

$$P_{k+1}(\pi,p)[\eta_{k+1}(p;i_{k+1})\in \underline{Y}_{k+1}|\eta_k(p;i_k)=\overline{y}_k, u_k=\overline{u}_k] = \hat{t}_k(\underline{Y}_{k+1}|\overline{y}_k,\overline{u}_k) \quad (22)$$

for $P_k(\pi,p)$ almost every $(\overline{y}_k,\overline{u}_k)$.†

(c) There exist lower semianalytic functions $\hat{g}_k:\hat{\Gamma}_k\to[-\infty,\infty]$ satisfying for every $p\in P(S)$, $\pi\in\Pi$, $k=0,\ldots,N-1$,

$$E[g(x_k,u_k)|\eta_k(p;i_k)=\overline{y}_k, u_k=\overline{u}_k] = \hat{g}_k(\overline{y}_k,\overline{u}_k) \quad (23)$$

for $P_k(\pi,p)$ almost every $(\overline{y}_k,\overline{u}_k)$, where the expectation is with respect to $P_k(\pi,p)$.

Condition (a) of Definition 10.6 guarantees that the control constraint set $U_k(i_k)$ can be recovered from $\eta_k(p;i_k)$. Indeed, from (15), (20), and (21), we have for any $p\in P(S)$, $i_k\in I_k$, $k=0,\ldots,N-1$,

$$U_k(i_k) = \hat{U}_k[\eta_k(p;i_k)]. \quad (24)$$

If $U_k(i_k)=C$ for every $i_k\in I_k$, $k=0,\ldots,N-1$, then condition (a) is satisfied with $\hat{\Gamma}_k = Y_k C$. This is the case of no control constraint. Condition (b) guarantees that the distribution of y_{k+1} depends only on the values of y_k and u_k. This is necessary in order for the variables y_k to form the states of a stochastic optimal control model of the type considered in Section 10.1. Condition (c) guarantees that the cost corresponding to a policy can be computed from the distributions induced on the (y_k,u_k) pairs.

We temporarily postpone discussion on the existence and the nature of particular statistics sufficient for control, and consider first a perfect state information model corresponding to model (ISI) and a given sufficient statistic.

Definition 10.7 Let the model (ISI) and a statistic sufficient for control $(\eta_0,\ldots,\eta_{N-1})$ be given. The *perfect state information stochastic optimal control model*, denoted by (PSI), consists of the following (we use the notation of Definitions 10.3 and 10.6):

Y_k, $k=0,\ldots,N-1$ *State spaces.*
C *Control space.*
$\hat{U}_k$, $k=0,\ldots,N-1$ *Control constraints.*
α *Discount factor.*
$\hat{g}_k$, $k=0,\ldots,N-1$ *One-stage cost functions.*
$\hat{t}_k$, $k=0,\ldots,N-2$ *State transition kernels.*
N *Horizon.*

† In this context "for $P_k(\pi,p)$ almost every $(\overline{y}_k,\overline{u}_k)$" means that the set $\{(x_0,z_0,u_0,\ldots,x_k,z_k,u_k)\in H_k|$ (22) holds when $\overline{y}_k=\eta_k(p;i_k),\overline{u}_k=u_k\}$ has $P_k(\pi,p)$-measure one.

Thus defined, (PSI) is a nonstationary stochastic optimal control model in the sense of Definition 10.1.† The definitions of policies and cost functions for (PSI) are given in Section 10.1. We will use ($\hat{}$) to denote these objects in (PSI). For example, $\hat{\Pi}'$ is the set of all (0-originating) policies and $\hat{\Pi}$ is the set of all Markov (0-originating) policies for (PSI). If $\hat{\pi} = (\hat{\mu}_0, \ldots, \hat{\mu}_{N-1})$ is a policy for (PSI), then by (24) and Proposition 7.44 the sequence

$$(\hat{\mu}_0[du_0|\eta_0(p;i_0)], \ldots, \hat{\mu}_{N-1}[du_{N-1}|\eta_0(p;i_0), u_0, \ldots, u_{N-2}, \eta_{N-1}(p;i_{N-1})]),$$

where

$$i_k = (z_0, u_0, \ldots, u_{k-1}, z_k), \qquad k = 0, \ldots, N-1, \tag{25}$$

is a policy for (ISI). We call this policy $\hat{\pi}$ also, and can regard $\hat{\Pi}'$ as a subset of Π in this sense. If $\hat{\pi}$ is a nonrandomized policy for (PSI), then it is also nonrandomized when considered as a policy for (ISI). We will see in Proposition 10.2 that $\hat{\pi}$ results in the same cost for both (PSI) and (ISI).

Define $\varphi: P(S) \to P(Y_0)$ by

$$\varphi(p)(\underline{Y}_0) = \int_S s_0(\{z_0|\eta_0(p;z_0) \in \underline{Y}_0\}|x_0)p(dx_0) \qquad \forall \underline{Y}_0 \in \mathscr{B}_{Y_0} \tag{26}$$

Thus defined, $\varphi(p)$ is the distribution of the initial state y_0 in (PSI) when the initial state x_0 in (ISI) has distribution p. By Corollary 7.26.1, for every $\underline{Y}_0 \in \mathscr{B}_{Y_0}$ the mapping

$$\psi_{\underline{Y}_0}(x_0, p) = s_0(\{z_0|\eta_0(p;z_0) \in \underline{Y}_0\}|x_0)$$

is Borel-measurable. Define a Borel-measurable stochastic kernel on S given $P(S)$ by $q(dx_0|p) = p(dx_0)$. Then (26) can be written as

$$\varphi(p)(\underline{Y}_0) = \int \psi_{\underline{Y}_0}(x_0, p)q(dx_0|p).$$

It follows from Propositions 7.26 and 7.29 that φ is Borel-measurable. For $p \in P(S)$, define the mapping $V_{p,k}: H_k \to Y_0C_0 \cdots Y_kC_k$ by

$$V_{p,k}(x_0, z_0, u_0, \ldots, x_k, z_k, u_k) = [\eta_0(p;i_0), u_0, \ldots, \eta_k(p;i_k), u_k], \tag{27}$$

where (25) holds. For $q \in P(Y_0)$ and $\hat{\pi} = (\hat{\mu}_0, \ldots, \hat{\mu}_{N-1}) \in \hat{\Pi}'$, there is a sequence of consistent probability measures $\hat{P}_k(\hat{\pi}, q)$ generated on $Y_0C_0 \cdots Y_kC_k$, $k = 0, \ldots, N-1$, defined on measurable rectangles by

$$\hat{P}_k(\hat{\pi}, q)(\underline{Y}_0\underline{C}_0 \cdots \underline{Y}_k\underline{C}_k) = \int_{\underline{Y}_0} \int_{\underline{C}_0} \cdots \int_{\underline{Y}_k} \hat{\mu}_k(\underline{C}_k|y_0, u_0, \ldots, u_{k-1}, y_k) \times \hat{t}_{k-1}(dy_k|y_{k-1}, u_{k-1}) \cdots \hat{\mu}_0(du_0|y_0)q(dy_0). \tag{28}$$

† The disturbance spaces, disturbance kernels, and system functions in (PSI) can be taken to be $W_k = Y_{k+1}$, $p_k(dw_k|y_k, u_k) = \hat{t}_k(dy_{k+1}|y_k, u_k)$, and $f_k(y_k, u_k, w_k) = w_k$, respectively.

For a Markov policy $\hat{\pi} \in \hat{\Pi}$, these objects are related to the probability measures $P_k(\hat{\pi}, p)$ defined by (16) in the following manner.

Lemma 10.1 Suppose $p \in P(S)$ and $\hat{\pi} \in \hat{\Pi}$. Then for $k = 0, \ldots, N-1$ and for every Borel set $B \subset Y_0 C_0 \cdots Y_k C_k$, we have

$$P_k(\hat{\pi}, p)[V_{p,k}^{-1}(B)] = \hat{P}_k[\hat{\pi}, \varphi(p)](B). \tag{29}$$

Proof It suffices to prove that if $\underline{Y}_0 \in \mathscr{B}_{Y_0}$, $\underline{C}_0 \in \mathscr{B}_{C_0}, \ldots, \underline{Y}_k \in \mathscr{B}_{Y_k}$, $\underline{C}_k \in \mathscr{B}_{C_k}$, then

$$\begin{aligned} &P_k(\hat{\pi}, p)(\{\eta_0(p; i_0) \in \underline{Y}_0, u_0 \in \underline{C}_0, \ldots, \eta_k(p; i_k) \in \underline{Y}_k, u_k \in \underline{C}_k\}) \\ &\quad = \hat{P}_k[\hat{\pi}, \varphi(p)](\underline{Y}_0 \underline{C}_0 \cdots \underline{Y}_k \underline{C}_k).^{\dagger} \end{aligned} \tag{30}$$

For $k = 0$, (30) follows from (16), (26), and (28). If (29) holds for some $k < N$, then using (16), (22), (28), and (29), we obtain

$$\begin{aligned} &P_{k+1}(\hat{\pi}, p)(\{\eta_0(p; i_0) \in \underline{Y}_0, u_0 \in \underline{C}_0, \ldots, \eta_{k+1}(p; i_{k+1}) \in \underline{Y}_{k+1}, u_{k+1} \in \underline{C}_{k+1}\}) \\ &\quad = \int_{\{\eta_0(p; i_0) \in \underline{Y}_0, u_0 \in \underline{C}_0, \ldots, \eta_k(p; i_k) \in \underline{Y}_k, u_k \in \underline{C}_k\}} \int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1} | y_{k+1}) \\ &\qquad \times \hat{t}_k(dy_{k+1} | \eta_k(p; i_k), u_k)\, dP_k(\hat{\pi}, p) \\ &\quad = \int_{\underline{Y}_0 \underline{C}_0 \ldots \underline{Y}_k \underline{C}_k} \int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1} | y_{k+1}) \hat{t}_k(dy_{k+1} | y_k, u_k)\, d\hat{P}_k[\hat{\pi}, \varphi(p)] \\ &\quad = \hat{P}_{k+1}[\hat{\pi}, \varphi(p)](\underline{Y}_0 \underline{C}_0 \cdots \underline{Y}_{k+1} \underline{C}_{k+1}). \qquad \text{Q.E.D.} \end{aligned}$$

As noted earlier, (PSI) is a model of the type considered in Section 10.1. The (F^+) and (F^-) conditions of Section 10.1, when specialized to the (PSI) model, will be denoted by $(\hat{F}^+)$ and $(\hat{F}^-)$, respectively. These conditions are not to be confused with the (F^+) and (F^-) conditions for the ISI model given in this section. In a particular problem it is often possible to see the relationship between these finiteness conditions on the two models. In the general case, the relationship is unclear. We point out, however, that if g is bounded below or above, then (F^+) or (F^-) is satisfied for (ISI), respectively, and given any statistic sufficient for control, the corresponding $\hat{g}_k$ can be chosen so as to be bounded below or above, respectively. If a particular result holds when we assume (F^+) on the (ISI) model and $(\hat{F}^+)$ on the (PSI) model, the notation $(F^+, \hat{F}^+)$ will appear. The notation $(F^-, \hat{F}^-)$ has a similar meaning.

† In this context, we define

$$\begin{aligned} &\{\eta_0(p; i_0) \in \underline{Y}_0, u_0 \in \underline{C}_0, \ldots, \eta_k(p; i_k) \in \underline{Y}_k, u_k \in \underline{C}_k\} \\ &\quad = \{(x_0, z_0, u_0, \ldots, x_k, z_k, u_k) | \eta_0(p; i_0) \in \underline{Y}_0, u_0 \in \underline{C}_0, \ldots, \eta_k(p; i_k) \in \underline{Y}_k, u_k \in \underline{C}_k\}, \end{aligned}$$

where $i_j = (z_0, u_0, \ldots, u_{j-1}, z_j)$. We will often use this notation to indicate a set which depends on functions of some or all of the components of a Cartesian product.

If $N = \infty$, we consider conditions (P), (N), and (D) for (ISI) and the corresponding conditions ($\hat{\text{P}}$), ($\hat{\text{N}}$), and ($\hat{\text{D}}$) for (PSI). In this case, however, if (P) holds for (ISI) and lower semianalytic functions $\hat{g}_k : \hat{\Gamma}_k \to [-\infty, \infty]$ satisfying (23) exist, there is no loss of generality in assuming that $\hat{g}_k \geq 0$ for every k, i.e., ($\hat{\text{P}}$) holds for (PSI). Likewise, if (N) or (D) holds for (ISI), we may assume without loss of generality that ($\hat{\text{N}}$) or ($\hat{\text{D}}$), respectively, holds for (PSI). As in the finite horizon case, we adopt the notation (P, $\hat{\text{P}}$), (N, $\hat{\text{N}}$), and (D, $\hat{\text{D}}$) to indicate which assumptions are sufficient for a result to hold.

From Section 10.1, we have that when ($\hat{\text{F}}^+$), ($\hat{\text{F}}^-$), ($\hat{\text{P}}$), ($\hat{\text{N}}$), or ($\hat{\text{D}}$) holds, then the (0-originating) *cost corresponding to a policy* $\hat{\pi}$ for (PSI) at $y \in Y_0$ is

$$\hat{J}_{N,\hat{\pi}}(y) = \sum_{k=0}^{N-1} \alpha^k \int_{Y_0 C_0 \cdots Y_k C_k} \hat{g}_k(y_k, u_k)\, d\hat{P}_k(\hat{\pi}, p_y), \tag{31}$$

where N may be infinite. The (0-originating) *optimal cost* for (PSI) at $y \in Y_0$ is

$$\hat{J}_N^*(y) = \inf_{\hat{\pi} \in \hat{\Pi}'} \hat{J}_{N,\hat{\pi}}(y). \tag{32}$$

The remainder of this section is devoted to establishing relations between costs, optimal costs, and optimal and nearly optimal policies for the (ISI) and (PSI) models.

Proposition 10.2 ($\text{F}^+, \hat{\text{F}}^+$)($\text{F}^-, \hat{\text{F}}^-$)(P, $\hat{\text{P}}$)(N, $\hat{\text{N}}$)(D, $\hat{\text{D}}$) For every $p \in P(S)$ and $\hat{\pi} \in \hat{\Pi}$, we have

$$J_{N,\hat{\pi}}(p) = \int_{Y_0} \hat{J}_{N,\hat{\pi}}(y_0)\varphi(p)(dy_0). \tag{33}$$

Proof From (31), (28), (23), (18), (19), and Lemma 10.1, we have

$$\begin{aligned}
\int_{Y_0} \hat{J}_{N,\hat{\pi}}(y)\varphi(p)(dy) &= \sum_{k=0}^{N-1} \alpha^k \int_{Y_0} \int_{Y_0 C_0 \cdots Y_k C_k} \hat{g}_k(y_k, u_k)\, d\hat{P}_k(\hat{\pi}, p_y)\varphi(p)(dy) \\
&= \sum_{k=0}^{N-1} \alpha^k \int_{Y_0 C_0 \cdots Y_k C_k} \hat{g}_k(y_k, u_k)\, d\hat{P}_k[\hat{\pi}, \varphi(p)] \\
&= \sum_{k=0}^{N-1} \alpha^k \int_{H_k} g(x_k, u_k)\, dP_k(\hat{\pi}, p) = J_{N,\hat{\pi}}(p),
\end{aligned}$$

where the (F^+) or (F^-) assumption is used to interchange integration and summation when $N < \infty$, and the monotone or bounded convergence theorem is used when $N = \infty$. Q.E.D.

Corollary 10.2.1 ($\text{F}^+, \hat{\text{F}}^+$)($\text{F}^-, \hat{\text{F}}^-$)(P, $\hat{\text{P}}$)(N, $\hat{\text{N}}$)(D, $\hat{\text{D}}$) For every $p \in P(S)$, we have

$$J_N^*(p) \leq \int_{Y_0} \hat{J}_N^*(y_0)\varphi(p)(dy_0). \tag{34}$$

Proof The function $\hat{J}_N^*(y_0)$ is lower semianalytic, so the integral in (34) is defined. From Proposition 10.2, we have

$$J_N^*(p) = \inf_{\pi \in \Pi} J_{N,\pi}(p) \leq \inf_{\hat{\pi} \in \hat{\Pi}} \int_{Y_0} \hat{J}_{N,\hat{\pi}}(y_0)\varphi(p)(dy_0),$$

so it suffices to show that

$$\inf_{\hat{\pi} \in \hat{\Pi}} \int_{Y_0} \hat{J}_{N,\hat{\pi}}(y_0)\varphi(p)(dy_0) = \int_{Y_0} \hat{J}_N^*(y_0)\varphi(p)(dy_0). \tag{35}$$

This follows from Lemma 8.6 and Corollary 9.5.2. Q.E.D.

We wish now to establish a relationship similar to (33) between the optimal cost functions for (ISI) and (PSI). In light of Corollary 10.2.1, it suffices to show that given any policy for (ISI), a policy for (PSI) can be found which does at least as well. This is formalized in the next lemma, and the analog of (33) is given as part of Proposition 10.3.

Lemma 10.2 $(F^+, \hat{F}^+)(F^-, \hat{F}^-)(P, \hat{P})(N, \hat{N})(D, \hat{D})$ Given $p \in P(S)$ and $\pi \in \Pi$, there exists $\hat{\pi} \in \hat{\Pi}$ such that

$$J_{N,\pi}(p) = \int_{Y_0} \hat{J}_{N,\hat{\pi}}(y_0)\varphi(p)(dy_0). \tag{36}$$

Proof Let $p \in P(S)$ and $\pi = (\mu_0, \ldots, \mu_{N-1}) \in \Pi$ be given. For $k = 0, \ldots, N-1$, let $Q_k(\pi, p)$ be the probability measure on $Y_k C_k$ defined on measurable rectangles to be

$$Q_k(\pi, p)(\underline{Y}_k \underline{C}_k) = P_k(\pi, p)(\{\eta_k(p; i_k) \in \underline{Y}_k, u_k \in \underline{C}_k\}). \tag{37}$$

There exists a Borel-measurable stochastic kernel $\hat{\mu}_k(du_k | y_k)$ on C_k given Y_k such that for every Borel set $B \subset Y_k C_k$ we have

$$Q_k(\pi, p)(B) = \int_{Y_k C_k} \hat{\mu}_k(B_{y_k} | y_k)\, dQ_k(\pi, p). \tag{38}$$

In particular,

$$\begin{aligned} 1 &= P_k(\pi, p)(\{(i_k, u_k) \in \Gamma_k\}) \\ &= P_k(\pi, p)(\{[\eta_k(p; i_k), u_k] \in \hat{\Gamma}_k\}) \\ &= Q_k(\pi, p)(\hat{\Gamma}_k) = \int_{Y_k C_k} \hat{\mu}_k(\hat{U}_k(y_k) | y_k)\, dQ_k(\pi, p), \end{aligned}$$

so, altering $\hat{\mu}_k(du_k | y_k)$ on a set of measure zero if necessary, we may assume that (38) holds and $\hat{\mu}_k(\hat{U}_k(y_k) | y_k) = 1$ for every $y_k \in Y_k$. Let $\hat{\pi} = (\hat{\mu}_0, \ldots, \hat{\mu}_{N-1})$. Then $\hat{\pi}$ is a Markov policy for (PSI).

We show by induction that for $\underline{Y}_k \in \mathscr{B}_{Y_k}$, $\underline{C}_k \in \mathscr{B}_C$, $k = 0, \ldots, N-1$,

$$Q_k(\pi, p)(\underline{Y}_k \underline{C}_k) = \hat{P}_k[\hat{\pi}, \varphi(p)](\{y_k \in \underline{Y}_k, u_k \in \underline{C}_k\}). \tag{39}$$

We see from (26) and (37) that the marginal of $Q_0(\pi, p)$ on Y_0 is $\varphi(p)$. Equation (39) for $k = 0$ follows from (28) and (38). Assume that (39) holds for k. From (38), (37), (22), and the induction hypothesis, we have

$$\begin{aligned}
&Q_{k+1}(\pi, p)(Y_{k+1}\underline{C}_{k+1}) \\
&\quad = \int_{\underline{Y}_{k+1}C_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1})\, dQ_{k+1}(\pi, p) \\
&\quad = \int_{\{\eta_{k+1}(p;\, i_{k+1}) \in \underline{Y}_{k+1}\}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|\eta_{k+1}(p; i_{k+1}))\, dP_{k+1}(\pi, p) \\
&\quad = \int_{H_k} \int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1}) \hat{t}_k(dy_{k+1}|\eta_k(p; i_k), u_k)\, dP_k(\pi, p) \\
&\quad = \int_{Y_k C_k} \int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1}) \hat{t}_k(dy_{k+1}|y_k, u_k)\, dQ_k(\pi, p) \\
&\quad = \int_{Y_0 C_0 \cdots Y_k C_k} \int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1}) \hat{t}_k(dy_{k+1}|y_k, u_k)\, d\hat{P}_k[\hat{\pi}, \varphi(p)] \\
&\quad = P_{k+1}[\hat{\pi}, \varphi(p)](\{y_{k+1} \in \underline{Y}_{k+1}, u_k \in \underline{C}_{k+1}\}).
\end{aligned}$$

Taken together, (37) and (39) imply that for $\underline{Y}_k \in \mathscr{B}_{Y_k}$, $\underline{C}_k \in \mathscr{B}_C$, $k = 0, \ldots, N-1$, we have

$$P_k(\pi, p)(\{\eta_k(p; i_k) \in \underline{Y}_k, u_k \in \underline{C}_k\}) = \hat{P}_k[\hat{\pi}, \varphi(p)](\{y_k \in \underline{Y}_k, u_k \in \underline{C}_k\}). \quad (40)$$

If (40) is used in place of Lemma 10.1, the proof of Proposition 10.2 can now be used to prove (36). Q.E.D.

Definition 10.8 Given $q \in P(Y_0)$ and $\varepsilon > 0$, a policy $\hat{\pi} \in \hat{\Pi}'$ is said to be *weakly q-ε-optimal* if

$$\int_{Y_0} \hat{J}_{N,\hat{\pi}}(y_0) q(dy_0) \leq \begin{cases} \int_{Y_0} \hat{J}_N^*(y_0) q(dy_0) + \varepsilon & \text{if } \int_{Y_0} \hat{J}_N^*(y_0) q(dy_0) > -\infty, \\ -1/\varepsilon & \text{if } \int_{Y_0} \hat{J}_N^*(y_0) q(dy_0) = -\infty. \end{cases}$$

The policy $\hat{\pi}$ is said to be *q-optimal* if $q(\{y_0 \in Y_0 | \hat{J}_{N,\hat{\pi}}(y_0) = \hat{J}_N^*(y_0)\}) = 1$.

Equation (35) shows that given any $p \in P(S)$ and $\varepsilon > 0$, a weakly $\varphi(p)$-ε-optimal Markov policy exists. The next proposition shows that such a policy is ε-optimal at p when considered as a policy in (ISI).

Proposition 10.3 $(F^+, \hat{F}^+)(F^-, \hat{F}^-)(P, \hat{P})(N, \hat{N})(D, \hat{D})$ We have

$$J_N^*(p) = \int_{Y_0} \hat{J}_N^*(y_0) \varphi(p)(dy_0) \qquad \forall p \in P(S). \quad (41)$$

Furthermore, if $\hat{\pi}$ is optimal, $\varphi(p)$-optimal, or weakly $\varphi(p)$-ε-optimal for (PSI), then $\hat{\pi}$ is optimal, optimal at p, or ε-optimal at p, respectively, for (ISI). If $\hat{\pi}$ is ε-optimal for (PSI) and $(F^+, \hat{F}^+)$, $(P, \hat{P})$, or $(D, \hat{D})$ holds, then $\hat{\pi}$ is also ε-optimal for (ISI).

Proof Equation (41) follows from Corollary 10.2.1 and Lemma 10.2. Let $\hat{\pi}$ be ε-optimal for (PSI). It is clear that under $(P, \hat{P})$ and $(D, \hat{D})$, we have

$$\hat{J}_N^*(y_0) > -\infty \qquad \forall y_0 \in Y_0, \tag{42}$$

so

$$\hat{J}_{N,\hat{\pi}}(y_0) \leq \hat{J}_N^*(y_0) + \varepsilon \qquad \forall y_0 \in Y_0. \tag{43}$$

Under $(F^+, \hat{F}^+)$, (42) follows from Lemma 8.3 and Proposition 8.2, so again (43) holds. We have from (41) and Proposition 10.2 that

$$\begin{aligned} J_{N,\hat{\pi}}(p) &= \int_{Y_0} \hat{J}_{N,\hat{\pi}}(y_0)\varphi(p)(dy_0) \\ &\leq \int_{Y_0} \hat{J}_N^*(y_0)\varphi(p)(dy_0) + \varepsilon \\ &= J_N^*(p) + \varepsilon \qquad \forall p \in P(S), \end{aligned}$$

so $\hat{\pi}$ is ε-optimal for (ISI). The remainder of the proposition follows from (41) and Proposition 10.2. Q.E.D.

We shall show shortly that a statistic sufficient for control always exists, and indeed, in many cases it can be chosen so that (PSI) is stationary. The existence of such a statistic for (ISI) and the consequent existence of the corresponding model (PSI) enable us to utilize the results of Chapters 8 and 9. For example, we have the following proposition.

Proposition 10.4 $(F^+, \hat{F}^+)(F^-, \hat{F}^-)(P, \hat{P})(N, \hat{N})(D, \hat{D})$ If $(\eta_0, \ldots, \eta_{N-1})$ is a statistic sufficient for control for (ISI), then for every $\varepsilon > 0$, there exists an ε-optimal nonrandomized policy for (ISI) which depends on $i_k = (z_0, u_0, \ldots, u_{k-1}, z_k)$ only through $\eta_k(p; i_k)$, i.e., has the form

$$\pi = (\mu_0[p; \eta_0(p; i_0)], \ldots, \mu_{N-1}[p; \eta_{N-1}(p; i_{N-1})]). \tag{44}$$

Under $(F^+, \hat{F}^+)$, $(P, \hat{P})$, or $(D, \hat{D})$, we may choose this ε-optimal policy to have the simpler form

$$\hat{\pi} = (\hat{\mu}_0[\eta_0(p; i_0)], \ldots, \hat{\mu}_{N-1}[\eta_{N-1}(p; i_{N-1})]). \tag{45}$$

Proof Under $(F^+, \hat{F}^+)$, $(P, \hat{P})$, or $(D, \hat{D})$, there exists an ε-optimal, nonrandomized, Markov policy $\hat{\pi} = (\hat{\mu}_0, \ldots, \hat{\mu}_{N-1})$ for (PSI) (Propositions 8.3 and 9.19). This policy $\hat{\pi}$ is ε-optimal for (ISI) by Proposition 10.3, and the second part of the proposition is proved.

Assume $(F^-, \hat{F}^-)$ holds and let $\{\varepsilon_n\}$ be a sequence of positive numbers with $\sum_{n=1}^{\infty} \varepsilon_n < \infty$ and $\varepsilon_n \downarrow 0$. Let $\hat{\pi}_n = (\hat{\mu}_0^n, \ldots, \hat{\mu}_{N-1}^n)$ be a sequence of nonrandomized Markov policies for (PSI) exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality (Proposition 8.4). By Proposition 10.2 and the $(F^-, \hat{F}^-)$

assumption, we have

$$\int_{Y_0} \hat{J}_{N,\hat{\pi}_n}(y_0)\varphi(p)(dy_0) = J_{N,\hat{\pi}_n}(p) < \infty \qquad \forall p \in P(S).$$

Since

$$\hat{J}_{N,\hat{\pi}_n}(y_0) + \sum_{k=n+1}^{\infty} \varepsilon_k \downarrow \hat{J}_N^*(y_0) \qquad \forall y_0 \in Y_0,$$

we have

$$\lim_{n\to\infty} \int_{Y_0} \hat{J}_{N,\hat{\pi}_n}(y_0)\varphi(p)(dy_0) = \int_{Y_0} \hat{J}^*(y_0)\varphi(p)(dy_0) \qquad \forall p \in P(S).$$

Let $\varepsilon > 0$ be given and let $n(p)$ be the smallest positive integer n for which

$$\int_{Y_0} \hat{J}_{N,\hat{\pi}_n}(y_0)\varphi(p)(dy_0) \leq \begin{cases} \int_{Y_0} \hat{J}^*(y_0)\varphi(p)(dy_0) + \varepsilon \\ \qquad \text{if} \quad \int_{Y_0} \hat{J}^*(y_0)\varphi(p)(dy_0) > -\infty, \\ -1/\varepsilon \\ \qquad \text{if} \quad \int_{Y_0} \hat{J}^*(y_0)\varphi(p)(dy_0) = -\infty. \end{cases}$$

Define $\mu_k(p; y_k) = \hat{\mu}_k^{n(p)}(y_k)$, $k = 0, \ldots, N-1$. Then by Propositions 10.2 and 10.3, π given by (44) is an ε-optimal nonrandomized policy for (ISI).

Assume $(N, \hat{N})$ holds. Consider the nonstationary stochastic optimal control model (NPSI) for which the initial state space is $P(Y_0)$, the initial control space is a singleton set $\{u_0\}$, the initial cost function is $g_0(q, u_0) = 0$ for every $q \in P(Y_0)$, and the initial transition kernel is given by $t(dy_0|q, u_0) = q(dy_0)$ for every $q \in P(Y_0)$. For $k \geq 0$, the $(k+1)$st state and control spaces, control constraint, cost function, and transition kernel are Y_k, C, $\hat{U}_k$, $\hat{g}_k$, and $\hat{t}_k(dy_{k+1}|y_k, u_k)$ of (PSI), respectively. The discount factor is α and the horizon is infinite. By definition, the optimal cost for (NPSI) at $q \in P(Y_0)$ is

$$\inf_{\hat{\pi}\in\hat{\Pi}} \int_{Y_0} \hat{J}_{\hat{\pi}}(y_0)q(dy_0),$$

which, by Corollaries 9.1.1 and 9.5.2, is the same as

$$\int_{Y_0} \hat{J}^*(y_0)q(dy_0).$$

Now (NPSI) has a nonpositive one-stage cost function, so, by Proposition 9.20, for each $\varepsilon > 0$ there exists an ε-optimal, nonrandomized, semi-Markov policy

$$\bar{\pi} = (\bar{\mu}(q), \bar{\mu}_0(q; y_0), \bar{\mu}_1(q; y_1), \ldots).$$

For fixed $q \in P(Y_0)$, let $\hat{\pi}(q)$ be the policy for (PSI) given by

$$\hat{\pi}(q) = (\bar{\mu}_0(q; y_0), \bar{\mu}_1(q; y_1), \ldots).$$

Then

$$\int_{Y_0} \hat{J}_{\hat{\pi}(q)}(y_0)q(dy_0) \leq \begin{cases} \int_{Y_0} \hat{J}^*(y_0)q(dy_0) + \varepsilon & \text{if } \int_{Y_0} \hat{J}^*(y_0)q(dy_0) > -\infty, \\ -1/\varepsilon & \text{if } \int_{Y_0} \hat{J}^*(y_0)q(dy_0) = -\infty, \end{cases}$$

i.e., $\hat{\pi}(q)$ is weakly q-ε-optimal for (PSI). By Proposition 10.3, the policy π defined by (44), where $\mu_k(p; y_k) = \bar{\mu}(\varphi(p); y_k)$, is ε-optimal for (ISI). Q.E.D.

The other specific results which can be derived for (ISI) from Chapters 8 and 9 are obvious and shall not be exhaustively listed. We content ourselves with describing the dynamic programming algorithm over a finite horizon.

By Proposition 8.2, the dynamic programming algorithm has the following form under $(F^+, \hat{F}^+)$ or $(F^-, \hat{F}^-)$, where we assume for notational simplicity that (PSI) is stationary:

$$\hat{J}_0^*(y) = 0 \qquad \forall y \in Y, \tag{46}$$

$$\hat{J}_{k+1}^*(y) = \inf_{u \in \hat{U}(y)} \{\hat{g}(y, u) + \alpha \int \hat{J}_k^*(y')\hat{t}(dy'|y, u)\}, \qquad k = 0, \ldots, N-1. \tag{47}$$

If the infimum in (47) is achieved for every y and $k = 0, \ldots, N-1$, then there exist universally measurable functions $\hat{\mu}_k: Y \to C$ such that for every y and $k = 0, \ldots, N-1$, $\hat{\mu}_k(y) \in \hat{U}(y)$ and $\hat{\mu}_k(y)$ achieves the infimum in (47). Then $\hat{\pi} = (\hat{\mu}_0, \ldots, \hat{\mu}_{N-1})$ is optimal in (PSI) (Proposition 8.5), so $\hat{\pi}$ is optimal in (ISI) as well (Proposition 10.3).

If $(F^+, \hat{F}^+)$ holds and the infimum in (47) is not achieved for every y and $k = 0, \ldots, N-1$, the dynamic programming algorithm (46) and (47) can still be used in the manner of Proposition 8.3 to construct an ε-optimal, nonrandomized, Markov policy $\hat{\pi}$ for (PSI). We see from Proposition 10.3 that $\hat{\pi}$ is an ε-optimal policy for (ISI) as well.

In many cases, $\eta_{k+1}(p; i_{k+1})$ is a function of $\eta_k(p; i_k)$, u_k, and z_{k+1}. The computational procedure in such a case is to first construct $(\hat{\mu}_0, \ldots, \hat{\mu}_{N-1})$ via (46) and (47), then compute $y_0 = \eta_0(p; i_0)$ from the initial distribution and the initial observation, and apply control $u_0 = \hat{\mu}_0(y_0)$. Given y_k, u_k, and z_{k+1}, compute y_{k+1} and apply control $u_{k+1} = \hat{\mu}_{k+1}(y_{k+1})$, $k = 0, \ldots, N-2$. In this way the information contained in $(p; i_k)$ has been condensed into y_k. This condensation of information is the historical motivation for statistics sufficient for control.

10.3 Existence of Statistics Sufficient for Control

Turning to the question of the existence of a statistic sufficient for control, it is not surprising to discover that the sequence of identity mappings on $P(S)I_k$, $k = 0, \ldots, N-1$, is such an object (Proposition 10.6). Although this

represents no condensation of information, it is sufficient to justify our analysis thus far. We will show that if the constraint sets Γ_k are equal to $I_k C$, $k = 0, \ldots, N-1$, then the functions mapping $P(S)I_k$ into the distribution of x_k conditioned on $(p; i_k)$, $k = 0, \ldots, N-1$, constitute a statistic sufficient for control (Proposition 10.5). This statistic has the property that its value at the $(k+1)$st stage is a function of its value at the kth stage, u_k and z_{k+1} [see (52)], so it represents a genuine condensation of information. It also results in a stationary perfect state information model and, if the conditional distributions can be characterized by a finite set of parameters, it may result in significant computational simplification. This latter condition is the case, for example, if it is possible to show beforehand that all these distributions are Gaussian.

10.3.1 Filtering and the Conditional Distributions of the States

We discuss filtering with the aid of the following basic lemma.

Lemma 10.3 Consider the (ISI) model. There exist Borel-measurable stochastic kernels $r_0(dx_0|p; z_0)$ on S given $P(S)Z$ and $r(dx|p; u, z)$ on S given $P(S)CZ$ which satisfy

$$\int_{\underline{S}_0} s_0(\underline{Z}_0|x_0)p(dx_0) = \int_S \int_{\underline{Z}_0} r_0(\underline{S}_0|p; z_0)s_0(dz_0|x_0)p(dx_0)$$

$$\forall \underline{S}_0 \in \mathscr{B}_S, \quad \underline{Z}_0 \in \mathscr{B}_Z, \quad p \in P(S), \tag{48}$$

$$\int_{\underline{S}} s(\underline{Z}|u, x)p(dx) = \int_S \int_{\underline{Z}} r(\underline{S}|p; u, z)s(dz|u, x)p(dx)$$

$$\forall \underline{S} \in \mathscr{B}_S, \quad \underline{Z} \in \mathscr{B}_Z, \quad p \in P(S), \quad u \in C. \tag{49}$$

Proof For fixed $(p; u) \in P(S)C$, define a probability measure q on SZ by specifying its values on measurable rectangles to be (Proposition 7.28)

$$q(\underline{S}\underline{Z}|p; u) = \int_{\underline{S}} s(\underline{Z}|u, x)p(dx).$$

By Propositions 7.26 and 7.29, $q(d(x, z)|p; u)$ is a Borel-measurable stochastic kernel on SZ given $P(S)C$. By Corollary 7.27.1, this stochastic kernel can be decomposed into its marginal on Z given $P(S)C$ and a Borel-measurable stochastic kernel $r(dx|p; u, z)$ on S given $P(S)CZ$ such that (49) holds. The existence of $r_0(dx_0|p; z_0)$ is proved in a similar manner. Q.E.D.

It is customary to call p, the given distribution of x_0, the *a priori distribution* of the initial state. After z_0 is observed, the distribution is "up-dated", i.e., the distribution of x_0 conditioned on z_0 is computed. The up-dated distribution is called the *a posteriori distribution* and, as we will show in Lemma 10.4, is just $r_0(dx_0|p; z_0)$. At the kth stage, $k \geq 1$, we will have some a priori distribution p'_k of x_k based on $i_{k-1} = (z_0, u_0, \ldots, u_{k-2}, z_{k-1})$. Control

u_{k-1} is applied, some z_k is observed, and an a posteriori distribution of x_k conditioned on (i_{k-1}, u_{k-1}, z_k) is computed. We will show that this distribution is just $r(dx|p'_k; u_{k-1}, z_k)$. The process of passing from an a priori to an a posteriori distribution in this manner is called *filtering*, and it is formalized next.

Consider the function $\bar{f}: P(S)C \to P(S)$ defined by

$$\bar{f}(p, u)(\underline{S}) = \int t(\underline{S}|x, u)p(dx) \qquad \forall \underline{S} \in \mathscr{B}_S. \tag{50}$$

Equation (50) is called the *one-stage prediction equation.* If x_k has an a posteriori distribution p_k and the control u_k is chosen, then the a priori distribution of x_{k+1} is $\bar{f}(p_k, u_k)$. The mapping $\bar{f}$ is Borel-measurable (Propositions 7.26 and 7.29).

Given a sequence $i_k \in I_k$ such that $i_{k+1} = (i_k, u_k, z_{k+1})$, $k = 0, \ldots, N-2$, and given $p \in P(S)$, define recursively

$$p_0(p; i_0) = r_0(dx_0|p; z_0), \tag{51}$$

$$p_{k+1}(p; i_{k+1}) = r(dx|\bar{f}[p_k(p; i_k), u_k]; u_k, z_{k+1}), \qquad k = 0, \ldots, N-2. \tag{52}$$

Note that for each k, $p_k: P(S)I_k \to P(S)$ is Borel-measurable.

Equations (48)–(52) are called the *filtering equations* corresponding to the (ISI) model. For a given initial distribution and policy, they generate the conditional distribution of the state given the current information, as the following lemma shows,

Lemma 10.4 Let the model (ISI) be given. For any $p \in P(S)$, $\pi = (\mu_0, \ldots, \mu_{N-1}) \in \Pi$ and $\underline{S}_k \in \mathscr{B}_S$, we have

$$P_k(\pi, p)[x_k \in \underline{S}_k | i_k] = p_k(p; i_k)(\underline{S}_k) \tag{53}$$

for $P_k(\pi, p)$ almost every i_k, $k = 0, \ldots, N-1$.

Proof† We proceed by induction. For any $\underline{S}_0 \in \mathscr{B}_S$ and $\underline{Z}_0 \in \mathscr{B}_Z$, we have from (51), (16), and (48), that

$$\begin{aligned}
\int_{\{z_0 \in \underline{Z}_0\}} p_0(p; z_0)(\underline{S}_0)\, dP_0(\pi, p) &= \int_{\{z_0 \in \underline{Z}_0\}} r_0(\underline{S}_0|p; z_0)\, dP_0(\pi, p) \\
&= \int_S \int_{\underline{Z}_0} r_0(\underline{S}_0|p; z_0) s_0(dz_0|x_0) p(dx_0) \\
&= \int_{\underline{S}_0} s_0(\underline{Z}_0|x_0) p(dx_0) \\
&= P_0(\pi, p)(\{x_0 \in \underline{S}_0, z_0 \in \underline{Z}_0\}).
\end{aligned} \tag{54}$$

Equation (53) for $k = 0$ follows from (54) and the definition of conditional probability.

† In this and subsequent proofs, the reader may find the discussion concerning conditional expectations and probabilities at the beginning of Section 10.2 helpful.

Assume now that (53) holds for k. For any $\underline{I}_k \in \mathscr{B}_{I_k}$, $\underline{C}_k \in \mathscr{B}_C$, $\underline{Z}_{k+1} \in \mathscr{B}_Z$ and $\underline{S}_{k+1} \in \mathscr{B}_S$, we have from (16), the induction hypothesis, Fubini's theorem, (50), (52), and (49) that

$$
\begin{aligned}
&\int_{\{i_k \in \underline{I}_k,\, u_k \in \underline{C}_k,\, z_{k+1} \in \underline{Z}_{k+1}\}} p_{k+1}(p; i_k, u_k, z_{k+1})(\underline{S}_{k+1})\, dP_{k+1}(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p; i_k, z_{k+1})(\underline{S}_{k+1}) s(dz_{k+1} | u_k, x_{k+1}) \\
&\qquad \times t(dx_{k+1} | x_k, u_k) \mu_k(du_k | p; i_k)\, dP_k(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{S_k} \int_{\underline{C}_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p; i_k, u_k, z_{k+1})(\underline{S}_{k+1}) s(dz_{k+1} | u_k, x_{k+1}) \\
&\qquad \times t(dx_{k+1} | x_k, u_k) \mu_k(du_k | p; i_k) [p_k(p; i_k)(dx_k)]\, dP_k(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p; i_k, u_k, z_{k+1})(\underline{S}_{k+1}) s(dz_{k+1} | u_k, x_{k+1}) \\
&\qquad \times t(dx_{k+1} | x_k, u_k) [p_k(p; i_k)(dx_k)] \mu_k(du_k | p; i_k)\, dP_k(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} r(\underline{S}_{k+1} | \bar{f}[p_k(p; i_k), u_k]; u_k, z_{k+1}) \\
&\qquad \times s(dz_{k+1} | u_k, x_{k+1}) \bar{f}[p_k(p; i_k), u_k](dx_{k+1}) \mu_k(du_k | p; i_k)\, dP_k(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{\underline{C}_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} | u_k, x_{k+1}) \bar{f}[p_k(p; i_k), u_k](dx_{k+1}) \\
&\qquad \times \mu_k(du_k | p; i_k)\, dP_k(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} | u_k, x_{k+1}) t(dx_{k+1} | x_k, u_k) [p_k(p; i_k)(dx_k)] \\
&\qquad \times \mu_k(du_k | p; i_k)\, dP_k(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{S_k} \int_{\underline{C}_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} | u_k, x_{k+1}) t(dx_{k+1} | x_k, u_k) \mu_k(du_k | p; i_k) \\
&\qquad \times [p_k(p; i_k)(dx_k)]\, dP_k(\pi, p) \\
&\quad = \int_{\{i_k \in \underline{I}_k\}} \int_{\underline{C}_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} | u_k, x_{k+1}) t(dx_{k+1} | x_k, u_k) \mu_k(du_k | p; i_k)\, dP_k(\pi, p) \\
&\quad = P_{k+1}(\pi, p)(\{i_k \in \underline{I}_k, u_k \in \underline{C}_k, x_{k+1} \in \underline{S}_{k+1}, z_{k+1} \in \underline{Z}_{k+1}\}). \qquad (55)
\end{aligned}
$$

It follows from (55) and the definition of conditional probability that

$$P_{k+1}(\pi, p)[x_{k+1} \in \underline{S}_{k+1} | i_{k+1}] = p_{k+1}(p; i_{k+1})(\underline{S}_{k+1})$$

for $P_{k+1}(\pi, p)$ almost every i_k, and the induction step is completed. Q.E.D.

Proposition 10.5 Consider the (ISI) model and assume that $U_k(x) = C$ for every $x \in S$ and $k = 0, \ldots, N-1$. Then the sequence $[p_0(p; i_0), \ldots, p_{N-1}(p; i_{N-1})]$ defined by (51) and (52) is a statistic sufficient for control, and the resulting perfect state information model is stationary.

Proof Let Y_k in Definition 10.6 be $P(S)$, $k = 0, \ldots, N-1$. We have already seen that the mappings $p_k: P(S)I_k \to P(S)$ are Borel-measurable, so $(p_0, \ldots, p_{N-1})$ is a statistic. Condition (a) of Definition 10.6 is satisfied with $\hat{\Gamma}_k = P(S)C$, $k = 0, \ldots, N-1$.

For $y \in P(S)$, $u \in C$ and $\underline{Y} \in \mathscr{B}_{P(S)}$, define

$$\underline{Z}(y, u, \underline{Y}) = \{z \in Z | r[dx|\bar{f}(y, u); u, z] \in \underline{Y}\},$$

$$\hat{t}(\underline{Y}|y, u) = \int_S \int_S s[\underline{Z}(y, u, \underline{Y})|u, x']t(dx'|x, u)y(dx). \tag{56}$$

Note that $\underline{Z}(y, u, \underline{Y})$ is the (y, u)-section of the inverse image of $\underline{Y}$ under a Borel-measurable function. The stochastic kernel

$$\lambda(\underline{Z}|x, u) = \int_S s(\underline{Z}|u, x')t(dx'|x, u)$$

is Borel-measurable by Propositions 7.26 and 7.29, so the stochastic kernel

$$\Lambda(\underline{Z}|y, u) = \int_S \int_S s(\underline{Z}|u, x')t(dx'|x, u)y(dx) = \int_S \lambda(\underline{Z}|x, u)y(dx)$$

is Borel-measurable by the same propositions. It follows from Proposition 7.26 and Corollary 7.26.1 that $\hat{t}(dy'|y, u)$ is a Borel-measurable stochastic kernel on $P(S)$ given $P(S)C$. For $\pi \in \Pi$, $p \in P(S)$, $\underline{Y} \in \mathscr{B}_{P(S)}$ and $k = 0, \ldots, N-2$, we have from Lemma 10.4

$$\begin{aligned}
&P_{k+1}(\pi, p)[p_{k+1}(p; i_{k+1}) \in \underline{Y}|p_k(p; i_k) = \bar{y}_k, u_k = \bar{u}_k] \\
&\quad = P_{k+1}(\pi, p)[z_{k+1} \in \underline{Z}(\bar{y}_k, \bar{u}_k, \underline{Y})|p_k(p; i_k) = \bar{y}_k, u_k = \bar{u}_k] \\
&\quad = E\{P_{k+1}(\pi, p)[z_{k+1} \in \underline{Z}(\bar{y}_k, \bar{u}_k, \underline{Y})|i_k, u_k]|p_k(p; i_k) = \bar{y}_k, u_k = \bar{u}_k\} \\
&\quad = E\Big\{\int_S \int_S s[\underline{Z}(\bar{y}_k, \bar{u}_k, \underline{Y})|u_k, x_{k+1}]t(dx_{k+1}|x_k, u_k) \\
&\qquad \times [p_k(p; i_k)(dx_k)]\Big|p_k(p; i_k) = \bar{y}_k, u_k = \bar{u}_k\Big\} \\
&\quad = \hat{t}(\underline{Y}|\bar{y}_k, \bar{u}_k)
\end{aligned}$$

for $P_k(\pi, p)$ almost every $(\bar{y}_k, \bar{u}_k)$, where the expectations are with respect to $P_{k+1}(\pi, p)$. Thus (22) is satisfied.

For $\pi \in \Pi$, $p \in P(S)$, and $k = 0, \ldots, N-1$, we have from Lemma 10.4

$$\begin{aligned}
&E[g(x_k, u_k)|p_k(p; i_k) = \bar{y}_k, u_k = \bar{u}_k] \\
&\quad = E\{E[g(x_k, u_k)|i_k, u_k]|p_k(p; i_k) = \bar{y}_k, u_k = \bar{u}_k\} \\
&\quad = E\Big\{\int_S g(x_k, u_k)p_k(p; i_k)(dx_k)\Big|p_k(p; i_k) = \bar{y}_k, u_k = \bar{u}_k\Big\} \\
&\quad = \int_S g(x_k, \bar{u}_k)\bar{y}_k(dx_k)
\end{aligned} \tag{57}$$

for $P_k(\pi, p)$ almost every $(\overline{y}_k, \overline{u}_k)$, where the expectations are with respect to $P_k(\pi, p)$. The function $\hat{g}: P(S)C \to R^*$ defined by

$$\hat{g}(\overline{y}, \overline{u}) = \int_S g(x, \overline{u})\overline{y}(dx) \tag{58}$$

is lower semianalytic (Proposition 7.48), and, by (57), $\hat{g}$ satisfies (23). Q.E.D.

If the horizon is finite, then the transition kernel $\hat{t}$ and the one-stage cost function $\hat{g}$ defined by (56) and (58) can be substituted in the dynamic programming algorithm (46)–(47) to compute the optimal cost function $\hat{J}_N^*$ for (PSI). The optimal cost function J_N^* for (ISI) can then be determined from (41). If the horizon is infinite, in the limit the dynamic programming algorithm (46)–(47) yields $\hat{J}^*$ under ($\hat{\mathrm{N}}$) and ($\hat{\mathrm{D}}$) and under ($\hat{\mathrm{P}}$) in some cases (Propositions 9.14 and 9.17). The determination of J^* from $\hat{J}^*$ is again accomplished by using (41).

10.3.2 The Identity Mappings

Proposition 10.6 Let the model (ISI) be given. The sequence of identity mappings on $P(S)I_k$, $k = 0, \ldots, N-1$, is a statistic sufficient for control.

Proof Let Y_k in Definition 10.6 be $P(S)I_k$, $k = 0, \ldots, N-1$, and let η_k be the identity mapping on $P(S)I_k$. Then $(\eta_0, \ldots, \eta_{N-1})$ is a statistic. Condition (a) of Definition 10.6 is satisfied with $\hat{\Gamma}_k = \mathrm{P}(S)\Gamma_k$, $k = 0, \ldots, N-1$.

If $\underline{Y}_{k+1} \in \mathscr{B}_{P(S)I_{k+1}}$, $\overline{y}_k \in P(S)I_k$, and $\overline{u}_k \in C_k$, we adopt the notation

$$(\underline{Y}_{k+1})_{(\overline{y}_k, \overline{u}_k)} = \{z_{k+1} \in Z \mid (\overline{p}; \overline{z}_0, \overline{u}_0, \ldots, \overline{u}_{k-1}, \overline{z}_k, \overline{u}_k, z_{k+1}) \in \underline{Y}_{k+1}\},$$

where $\overline{y}_k = (\overline{p}; \overline{z}_0, \overline{u}_0, \ldots, \overline{u}_{k-1}, \overline{z}_k)$. Using this notation, we define for $k = 0, \ldots, N-2$ the stochastic kernel $\hat{t}_k(dy_{k+1} \mid \overline{y}_k, \overline{u}_k)$ on $P(S)I_{k+1}$ given $P(S)I_kC$ by

$$\hat{t}_k(\underline{Y}_{k+1} \mid \overline{y}_k, \overline{u}_k) = \int_{S_{k+1}} s[(\underline{Y}_{k+1})_{(\overline{y}_k, \overline{u}_k)} \mid \overline{u}_k, x_{k+1}] t(dx_{k+1} \mid x_k, \overline{u}_k) p_k(\overline{y}_k)(dx_k)$$
$$\forall \underline{Y}_{k+1} \in \mathscr{B}_{P(S)I_{k+1}}, \tag{59}$$

where $p_k(\overline{y}_k)$ is given by (51) and (52). By an argument similar to that used in Proposition 10.5, it can be shown that $\hat{t}_k$ is Borel-measurable. For $p \in P(S)$, $\pi \in \Pi$, $\underline{Y}_{k+1} \in \mathscr{B}_{P(S)I_{k+1}}$, and $k = 0, \ldots, N-2$, we have from Lemma 10.4

$$\begin{aligned} &P_{k+1}(\pi, p)[\eta_{k+1}(p; i_{k+1}) \in \underline{Y}_{k+1} \mid \eta_k(p; i_k) = \overline{y}_k, u_k = \overline{u}_k] \\ &\quad = P_{k+1}(\pi, p)[(\overline{y}_k, \overline{u}_k, z_{k+1}) \in \underline{Y}_{k+1}] \\ &\quad = \int_{S_{k+1}} s[(\underline{Y}_{k+1})_{(\overline{y}_k, \overline{u}_k)} \mid \overline{u}_k, x_{k+1}] t(dx_{k+1} \mid x_k, \overline{u}_k) p_k(\overline{y}_k)(dx_k) \\ &\quad = \hat{t}_k(\underline{Y}_{k+1} \mid \overline{y}_k, \overline{u}_k), \end{aligned}$$

for $P_k(\pi, p)$ almost every $(\overline{y}_k, \overline{u}_k)$, so (22) is satisfied.

For $k = 0, \ldots, N-1$, define $\hat{g}_k : P(S)I_kC \to R^*$ by

$$\hat{g}_k(\bar{y}_k, \bar{u}_k) = \int_{S_k} g(x_k, \bar{u}_k) p_k(\bar{y}_k)(dx_k). \tag{60}$$

By Proposition 7.48, $\hat{g}_k$ is lower semianalytic for each k. For $p \in P(S)$, $\pi \in \Pi$, and $k = 0, \ldots, N-1$, we have from Lemma 10.4

$$\begin{aligned} E[g(x_k, u_k) | \eta(p; i_k) = \bar{y}_k, u_k = \bar{u}_k] &= \int_{S_k} g(x_k, \bar{u}_k) p_k(\bar{y}_k)(dx_k) \\ &= \hat{g}_k(\bar{y}_k, \bar{u}_k) \end{aligned}$$

for $P_k(\pi, p)$ almost every $(\bar{y}_k, \bar{u}_k)$, where the expectation is with respect to $P_k(\pi, p)$, so (23) is satisfied. Q.E.D.

The transition kernels $\hat{t}_k$ and one-stage cost functions $\hat{g}_k$ defined by (59) and (60) can be used in the nonstationary version of the dynamic programming algorithm (46)–(47). See the discussion following Proposition 10.5.

Chapter 11

Miscellaneous

11.1 Limit-Measurable Policies

In this section we strengthen the results of Section 7.7 concerning universally measurable functions. In particular, we show that these results are still valid if limit-measurable functions (Definitions B.2 and B.3) are used in place of universally measurable functions. This allows us to replace all the results on the existence of universally measurable policies in Chapters 8 and 9 by stronger results on the existence of limit-measurable policies.

We now rework the main results of Section 7.7 with the aid of the concepts and results of Appendix B.

Proposition 11.1 Let X, Y, and Z be Borel spaces, $D \in \mathscr{L}_X$, and $E \in \mathscr{L}_Y$. Suppose $f: D \to Y$ and $g: E \to Z$ are limit-measurable and $f(D) \subset E$. Then the composition $g \circ f$ is limit-measurable.

Proof This follows from Corollary B.11.1. Q.E.D.

Corollary 11.1.1 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 limit-measurable if and only if f is limit-measurable.

Proof See the proof of Corollary 7.44.3. Q.E.D.

Proposition 11.2 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 limit-measurable.

(b) For every $B \in \mathscr{B}_Y$, the mapping $\lambda_B: X \to R$ defined by

$$\lambda_B(x) = q(B|x) \tag{1}$$

is limit-measurable.

(c) For every $Q \in \mathscr{L}_Y$, the mapping λ_Q of (1) is limit-measurable.

Proof We prove (a) $\Rightarrow$ (c) $\Rightarrow$ (b) $\Rightarrow$ (a). Suppose (a) holds and $Q \in \mathscr{L}_Y$. Now $\lambda_Q = \theta_Q \circ \gamma$, where $\gamma: X \to P(Y)$ is given by

$$\gamma(x) = q(dy|x) \tag{2}$$

and $\theta_Q: P(Y) \to R$ is given by

$$\theta_Q(q) = q(Q). \tag{3}$$

We have assumed that γ is limit-measurable, and θ_Q is limit-measurable by Proposition B.12. Therefore (c) holds.

It is clear that (c) $\Rightarrow$ (b). Suppose now that (b) holds. Then

$$\sigma\left[\bigcup_{B \in \mathscr{B}_Y} \lambda_B^{-1}(\mathscr{B}_R)\right] \subset \mathscr{L}_X.$$

Letting γ and θ_B be defined by (2) and (3), we have from Proposition 7.25

$$\begin{aligned}\gamma^{-1}[\mathscr{B}_{P(Y)}] &= \gamma^{-1}\left[\sigma\left(\bigcup_{B \in \mathscr{B}_Y} \theta_B^{-1}(\mathscr{B}_R)\right)\right] \\ &= \sigma\left[\bigcup_{B \in \mathscr{B}_Y} \gamma^{-1}(\theta_B^{-1}(\mathscr{B}_R))\right] = \sigma\left[\bigcup_{B \in \mathscr{B}_Y} \lambda_B^{-1}(\mathscr{B}_R)\right] \subset \mathscr{L}_X,\end{aligned}$$

so $q(dy|x)$ is limit-measurable. Q.E.D.

Proposition 11.3 Let X and Y be Borel spaces and let $f: XY \to R^*$ be limit-measurable. Let $q(dy|x)$ be a limit-measurable stochastic kernel on Y given X. Then the mapping $\lambda: X \to R^*$ defined by

$$\lambda(x) = \int f(x, y) q(dy|x)$$

is limit-measurable.

Proof The mapping $\delta(x) = p_x$ is continuous (Corollary 7.21.1), as is the mapping $\sigma: P(X)P(Y) \to P(XY)$ defined by $\sigma(p, q) = pq$, where pq is the

product measure (Lemma 7.12). Suppose $Q \in \mathscr{L}_{XY}$ and $f = \chi_Q$. For every $x \in X$,

$$\lambda(x) = [p_x q(dy|x)](Q) = \theta_Q(\sigma[\delta(x), \gamma(x)]), \tag{4}$$

where γ and θ_Q are given by (2) and (3). Since all the functions on the right-hand side of (4) are limit-measurable, λ is limit-measurable. It follows that λ is limit-measurable when f is a limit-measurable simple function. The extension to the general limit-measurable, extended real-valued function f is straightforward. Q.E.D.

Corollary 11.3.1 Let X be a Borel space and let $f: X \to R^*$ be limit-measurable. Then the function $\theta_f: P(X) \to R^*$ defined by

$$\theta_f(p) = \int f \, dp$$

is limit-measurable.

We have the following sharpened version of the selection theorem for lower semianalytic functions.

Proposition 11.4 Let X and Y be Borel spaces, $D \subset XY$ an analytic set, and $f: D \to R^*$ a lower semianalytic function. Define $f^*: \text{proj}_X(D) \to R^*$ by

$$f^*(x) = \inf_{y \in D_x} f(x, y).$$

The set

$$I = \{x \in \text{proj}_X(D) | \text{ for some } y_x \in D_x, f(x, y_x) = f^*(x)\}$$

is limit-measurable, and for every $\varepsilon > 0$ there exists a limit-measurable function $\varphi: \text{proj}_X(D) \to Y$ such that $\text{Gr}(\varphi) \subset D$ and for all $x \in \text{proj}_X(D)$

$$f[x, \varphi(x)] = f^*(x) \qquad \text{if} \quad x \in I,$$

$$f[x, \varphi(x)] \leq \begin{cases} f^*(x) + \varepsilon & \text{if} \quad x \notin I, \quad f^*(x) > -\infty, \\ -1/\varepsilon & \text{if} \quad x \notin I, \quad f^*(x) = -\infty. \end{cases}$$

Proof The proof is the same as in Proposition 7.50(b), except that at the points where Corollary 7.44.2 is invoked to say that the composition of analytically measurable functions is universally measurable, we use Proposition 11.1 to say that the composition is limit-measurable. Q.E.D.

By the remark following Corollary B.11.1, we see that *I and the selector obtained in Proposition 11.4 are in fact $\mathscr{L}_X^2$-measurable.* This remark further suggests that the constructions in Chapters 8 and 9 of optimal and ε-optimal

policies can be done more carefully by keeping track of the minimal $\mathscr{L}_S^x$ with respect to which policies and costs are measurable. We do this to some extent in the next section, but do not pursue this matter to any great length.

Propositions 11.1–11.4 are sufficient to allow us to replace every reference to a "(universally measurable) policy" in Chapters 8 and 9 by the words "limit-measurable policy." It does not matter which class of policies is considered when defining J_N^* and J^*; the proof of Proposition 8.1 together with Proposition 11.5 given below can be used to show that these functions are determined by the analytically measurable Markov policies alone. Corollary 11.1.1 tells us that the nonrandomized limit-measurable policies are just the set of sequences of limit-measurable functions from state to control which satisfy the control constraint (cf. Definition 8.2). This fact and Proposition 11.2 are needed for the proof of the limit-measurable counterpart of Lemma 8.2. From Proposition 11.3 we can deduce that the cost corresponding to a limit-measurable policy is limit-measurable (cf. Definitions 8.3 and 9.3). This fact was used, for example, in proving that under (F^-) a nonrandomized, semi-Markov, ε-optimal policy exists (Proposition 8.3). Proposition 11.4 allows limit-measurable ε-optimal and optimal selection. The ε-optimal selection property for universally measurable functions is used in practically every proof in Chapters 8 and 9. The exact selection property is used in showing the existence under certain conditions of optimal policies (Propositions 8.5, 9.19, and 9.20).

11.2 Analytically Measurable Policies

Some of the existence results of Chapters 8 and 9 can be sharpened to state the existence of ε-optimal analytically measurable policies. This is due to Proposition 7.50(a) and the following propositions. Proposition 11.5 is the analog of Corollary 7.44.3 for universally measurable policies and of Corollary 11.1.1 for limit-measurable policies.

Proposition 11.5 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 analytically measurable if and only if f is analytically measurable.

Proof We sharpen the proof of Corollary 7.44.3. Let $\gamma(x) = q(dy|x)$ and $\delta(y) = p_y$, so that $\gamma = \delta \circ f$ and $f = \delta^{-1} \circ \gamma$. Now δ is a homeomorphism from Y to $\bar{Y} = \{p_y | y \in Y\}$, so δ and $\delta^{-1}: \bar{Y} \to Y$ are both Borel-measurable. If f is analytically measurable and $C \in \mathscr{B}_{P(Y)}$, then

$$\gamma^{-1}(C) = f^{-1}[\delta^{-1}(C)] \in \mathscr{A}_X$$

because $\delta^{-1}(C) \in \mathscr{B}_Y$. If γ is analytically measurable and $B \in \mathscr{B}_Y$, then

$$f^{-1}(B) = \gamma^{-1}[\delta(B)] \in \mathscr{A}_X$$

because $\delta(B) \in \mathscr{B}_{P(Y)}$. Q.E.D.

Proposition 11.6 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 analytically measurable.
(b) For every $B \in \mathscr{B}_Y$, the mapping $\lambda_B: X \to R$ defined by

$$\lambda_B(x) = q(B|x) \tag{5}$$

is analytically measurable.

Proof Assume (a) holds and define $\gamma(x) = q(dy|x)$. Then for $B \in \mathscr{B}_Y$, $C \in \mathscr{B}_R$, and $\theta_B: P(Y) \to R$ defined by (3), we have

$$\lambda_B^{-1}(C) = \gamma^{-1}[\theta_B^{-1}(C)] \in \mathscr{A}_X$$

because $\theta_B^{-1}(C) \in \mathscr{B}_{P(Y)}$ (Proposition 7.25). Therefore (b) holds.

If (b) holds, we can show that (a) holds by the same argument used in the proof of Proposition 11.2. Q.E.D.

We know from Corollary B.11.1 that the composition of analytically measurable functions need not be analytically measurable, so the cost corresponding to an analytically measurable policy for a stochastic optimal control model may not be analytically measurable. To see this, just write out explicitly the cost corresponding to a two-stage, nonrandomized, Markov, analytically measurable policy (cf. Definition 8.3).

A review of Chapters 8 and 9 shows the following. Proposition 8.3 is still valid if the word "policy" is replaced by "analytically measurable policy," except that under (F^-) an analytically measurable, nonrandomized, semi-Markov, ε-optimal policy is not guranteed to exist. However, an analytically measurable nonrandomized ε-optimal policy can be shown to exist if $g \leq 0$ [B12]. The proof of the existence of a sequence of nonrandomized Markov policies exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality (Proposition 8.4) breaks down at the point where we assume that a sequence of one-stage policies $\{\mu_0^n\}$ exists for which

$$T_{\mu_0^n}(J_0) \leq T_{\mu_0^{n-1}}(J_0).$$

This occurs because $T_{\mu_0^{n-1}}(J_0)$ may not be analytically measurable. In the first sentence of Proposition 9.19, the word "policy" can be replaced by "analytically measurable policy." The ε-optimal part of Proposition 9.20 depends on the (F^-) part of Proposition 8.3, so it cannot be strengthened in

this way. Under assumption (N), an analytically measurable, nonrandomized, ε-optimal policy can be shown to exist [B12], but it is unknown whether this policy can be taken to be semi-Markov. The results of Chapters 8 and 9 relating to existence of universally measurable optimal policies depend on the exact selection property of Proposition 7.50(b). Since this property is not available for analytically measurable functions, we cannot use the same arguments to infer existence of optimal analytically measurable policies.

11.3 Models with Multiplicative Cost

In this section we revisit the stochastic optimal control model with a multiplicative cost functional first encountered in Section 2.3.4. We pose the finite horizon model in Borel spaces and state the results which are obtainable by casting this Borel space model in the generalized framework of Chapter 6. This does not permit a thorough treatment of the type already given to the model with additive cost in Chapters 8 and 9, but it does yield some useful results and illustrates how the generalized abstract model of Chapter 6 can be applied. The reader can, of course, use the mathematical theory of Chapter 7 to analyze the model with multiplicative cost directly under conditions more general than those given here.

We set up the *Borel model with multiplicative cost*. Let the state space S, the control space C, and the disturbance space W be Borel spaces. Let the control constraint U mapping S into the set of nonempty subsets of C be such that

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

is analytic. Let the disturbance kernel $p(dw|x,u)$ and the system function $f: SCW \to S$ be Borel-measurable. Let the one-stage cost function g be Borel-measurable, and assume that there exists a $b \in R$ such that $0 \leq g(x,u,w) \leq b$ for all $x \in S$, $u \in U(x)$, $w \in W$. Let the horizon N be a positive integer.

In the framework of Section 6.1, we define $\tilde{F}$ to be the set of extended real-valued, universally measurable functions on S and F^* to be the set of functions in $\tilde{F}$ which are lower semianalytic. We let $\tilde{M}$ be the set of universally measurable functions from S to C with graph in Γ. Define $H: SC\tilde{F} \to [0, \infty]$ by

$$H(x,u,J) = \int_W g(x,u,w) J[f(x,u,w)] p(dw|x,u),$$

where we define $0 \cdot \infty = \infty \cdot 0 = 0 \cdot (-\infty) = (-\infty) \cdot 0 = 0$. We take $J_0: S \to R^*$ to be identically one. Then Assumptions A.1–A.4, $\tilde{\text{F}}$.2, and the Exact Selection Assumption of Section 6.1 hold. (Assumption A.2 follows from Lemma 7.30(4) and Propositions 7.47 and 7.48. Assumption A.4 follows from Proposition

7.50.) From Propositions 6.1(a), 6.2(a), and 6.3 we have the following results, where the notation of Section 6.1 is used.

Proposition 11.7 In the finite horizon Borel model with multiplicative cost, we have

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

and for every $\varepsilon > 0$ there exists an N-stage ε-optimal (Markov) policy. A policy $\pi^* = (\mu_0^*, \ldots, \mu_{N-1}^*)$ is uniformly N-stage optimal if and only if $(T_{\mu_k^*} T^{N-k-1})(J_0) = T^{N-k}(J_0)$, $k = 0, \ldots, N-1$, and such a policy exists 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 $k = 0, \ldots, N-1$. A sufficient condition for this infimum to be attained is for the set

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

to be compact for each $x \in S$, $\lambda \in R$, and $k = 0, \ldots, N-1$.