5

Models with Restricted Policies

Contents

In this chapter, we discuss variants of some of the models of the preceding chapters, where there are restrictions on the set of policies. In particular, policies may be selected from a strict subset $\overline{\mathcal{M}}$ of the set of functions $\mu : X \mapsto U$ with $\mu(x) \in U(x)$ for all $x \in X$. One potential use of such a restriction arises when $\overline{\mathcal{M}}$ consists of specially structured policies that result in convenient characterization and computation. Classical examples of situations of this type are linear policies in linear-quadratic optimal control problems, (s, S) policies in inventory control, and various threshold policies in queueing and scheduling problems (see e.g., [Ber05a], [Ber12a], [Put94]). In this context, the main focus of the analysis is to show that attention can be confined to the policies $\mu \in \overline{\mathcal{M}}$ and their associated mappings T_μ. If this can be shown, then the analytical solution of Bellman's equation may be enhanced, and the computational solution using for example policy iteration that is restricted within $\overline{\mathcal{M}}$ may be facilitated (cf. Section 4.4.3, which deals with linear-quadratic optimal control).

Another major use arises in mathematically rigorous probabilistic treatments of stochastic optimal control problems, and relates to the need for μ to be measurable in an appropriate sense. We take this as the starting point and motivation for the development of a general theory of restricted models in the next section, and we develop this theory in subsequent sections for finite horizon and for contractive models. In the last section of this chapter, we return to the treatment of measurability issues using the theory developed in the earlier sections.

5.1 A FRAMEWORK FOR RESTRICTED POLICIES

As a motivating example, let us consider the mapping H of the stochastic optimal control Example 1.2.1,

$$H(x, u, J) = E\big\{g(x, u, w) + \alpha J\big(f(x, u, w)\big)\big\}. \tag{5.1}$$

We have considered so far the case where w is a discrete random variable, taking a finite or a countably infinite set of values, so that the mapping T_μ is defined for all $\mu \in \mathcal{M}$ in terms of a summation. In the general case, however, where w is a continuous random variable, H is well-defined only if the expected value over w is well-defined. For this it is necessary to introduce an appropriate probability space for w, and for g and f to be appropriately measurable.

Most importantly, we cannot simply take J to belong to $E(X)$, the space of extended real-valued functions over X. We must restrict J to a subset $\overline{E}(X) \subset E(X)$, consisting of appropriately measurable functions. In addition, to define T_μ as a function from $\overline{E}(X)$ to $\overline{E}(X)$ by

$$(T_\mu J)(x) = H\big(x, \mu(x), J\big)$$

or equivalently

$$(T_\mu J)(x) = E\big\{g(x,\mu(x),w) + \alpha J\big(f(x,\mu(x),w)\big)\big\},$$

μ must belong to a subclass $\overline{\mathcal{M}}$ of appropriately measurable functions from $\mathcal{M}$. Appendix C provides a further illustration of the measurability issues through a simple two-stage example, and also introduces some of the terminology and background on measure theoretic issues, based on Borel measurability, which we will use in this chapter (the monograph [BeS78] contains an extensive account of this material).

The preceding discussion indicates that to define a restricted policies model, we need at least two basic subsets of functions and policies:

(a) A *restricted set of functions* $\overline{E}(X) \subset E(X)$.

(b) A *restricted set of policies* $\overline{\mathcal{M}} \subset \mathcal{M}$, such that

$$T_\mu J \in \overline{E}(X), \qquad \forall\ J \in \overline{E}(X),\ \mu \in \overline{\mathcal{M}},$$

where T_μ is given by

$$(T_\mu J)(x) = H\big(x,\mu(x),J\big), \qquad \forall\ x \in X.$$

To conduct an analysis similar to the one of earlier chapters, we also need to concern ourselves with the corresponding mapping T. To this end, we assume, without loss of generality, that

$$U(x) = \big\{\mu(x) \mid \mu \in \overline{\mathcal{M}}\big\}, \qquad \forall\ x \in X,$$

so that the mapping T can be interchangeably defined as

$$(TJ)(x) = \inf_{\mu \in \overline{\mathcal{M}}} (T_\mu J)(x), \qquad \forall\ J \in \overline{E}(X),\ x \in X,$$

or as

$$(TJ)(x) = \inf_{u \in U(x)} H(x,u,J), \qquad \forall\ J \in \overline{E}(X),\ x \in X. \tag{5.2}$$

An issue to contend with here is whether the infimum in the definition of TJ is attained, exactly or within a small tolerance $\epsilon > 0$, by a policy in $\overline{\mathcal{M}}$ (simultaneously for all x), i.e., whether there exists $\mu \in \overline{\mathcal{M}}$ such that $(T_\mu J)(x)$ is close to $(TJ)(x)$ uniformly for all $x \in X$. This issue turns out to be quite delicate in the context of stochastic optimal control problems, as we will discuss shortly. We first consider the case where the infimum is exactly attained, and then address the more complex case where it is not.

Models Admitting Exact Selection

Let us assume that $\overline{E}(X)$ and $\overline{\mathcal{M}}$ have been defined as described above. Consider the case where there exists $\mu \in \overline{\mathcal{M}}$ such that the infimum in Eq. (5.2) is attained at $\mu(x)$ for all $x \in X$, i.e., for all $J \in \overline{E}(X)$ there exists a $\mu \in \overline{\mathcal{M}}$ such that

$$T_\mu J = TJ. \tag{5.3}$$

Otherwise stated, *for all $J \in \overline{E}(X)$, the minimization in Eq. (5.2) admits an exact selector μ from within $\overline{\mathcal{M}}$*. Then assuming also that T and T_μ, $\mu \in \overline{\mathcal{M}}$, preserve membership in $\overline{E}(X)$, i.e.,

$$T_\mu J \in \overline{E}(X), \quad TJ \in \overline{E}(X), \qquad \forall\ J \in \overline{E}(X),\ \mu \in \overline{\mathcal{M}}, \tag{5.4}$$

a large portion of the analysis of the preceding chapters carries through verbatim, and much of the remainder can be extended with minimal modifications.

In particular, in the finite horizon problems of Chapter 4, under this condition, the condition $\bar{J} \in \overline{E}(X)$, and the monotonicity assumption

$$H(x,u,J) \leq H(x,u,J'), \qquad \forall\ J, J' \in \overline{E}(X),\ x \in X,\ u \in U(x), \tag{5.5}$$

we have $J_N^* = T^N \bar{J}$ and that there exists an N-stage optimal policy. Such a policy can be obtained via the DP algorithm that starts with the terminal cost function $\bar{J}$, and sequentially computes $T\bar{J}, T^2\bar{J}, \ldots, T^N\bar{J}$, and corresponding $\mu_{N-1}^*, \mu_{N-2}^*, \ldots, \mu_0^* \in \overline{\mathcal{M}}$ such that

$$T_{\mu_k^*} T^{N-k-1} \bar{J} = T^{N-k} \bar{J}, \qquad k = 0, \ldots, N-1, \tag{5.6}$$

(cf. the discussion in the early part of Section 4.2).

To extend the analysis of the contractive models of Chapter 2, under Eqs. (5.3) and (5.4), we need to assume that $\overline{E}(X)$ is a *closed* subset of $B(X)$, the space of functions $J : X \mapsto \Re$ that are bounded with respect to a weighted sup-norm. This is necessary so that the fixed point theorems of Appendix B apply. We also need to assume that the mappings T_μ are contractions for all $\mu \in \overline{\mathcal{M}}$ with respect to a common weighted sup-norm. Then the relevant portion of the analysis of Chapter 2 carries through with hardly any change. The analysis of the semicontractive and infinite horizon noncontractive models of Chapters 3 and 4, also admit a similar treatment, under the exact selection assumption (5.3) and Eq. (5.4).

Generally, when the exact selection property (5.3) holds in the context of the stochastic optimal control example where H is defined by Eq. (5.1), there are few complications in providing a rigorous mathematical treatment, even when w is a continuous random variable. Typically X, U, and W are taken to be Borel spaces, $\overline{E}(X)$ and $\overline{\mathcal{M}}$ are chosen to be the spaces of Borel measurable functions from X to $\Re^*$, and from X to U, respectively. Also g, f, and the probability space of w must satisfy certain Borel measurability conditions (see Appendix C).

Models Without Exact Selection

When the exact selection property (5.3) may not hold, to conduct any kind of meaningful analysis, it is necessary to adopt a restriction framework for policies and functions, which guarantees that TJ can be approximated by $T_\mu J$, with appropriate choice of μ. To this end, a seemingly natural assumption would be that given $J \in \overline{E}(X)$ and $\epsilon > 0$, *there exists an ϵ-optimal selector*, that is, a $\mu_\epsilon \in \overline{\mathcal{M}}$ such that

$$(T_{\mu_\epsilon} J)(x) \leq \begin{cases} (TJ)(x) + \epsilon & \text{if } (TJ)(x) > -\infty, \\ -(1/\epsilon) & \text{if } (TJ)(x) = -\infty, \end{cases} \qquad \forall\ x \in X. \tag{5.7}$$

However, in the Borel space model noted earlier and described in Appendix C, there is a serious difficulty: *if $\overline{E}(X)$ and $\overline{\mathcal{M}}$ are the spaces of Borel measurable functions from X to $\Re^*$, and from X to U, respectively, there need not exist an ϵ-optimal selector*. For this reason, Borel measurability of cost functions and policies is not the most appropriate probabilistic framework for stochastic optimal control problems.† Instead, in the most general framework for bypassing this difficulty, it is necessary to consider a different kind of measurability, which is described in Appendix C. In this framework:

(a) $\overline{E}(X)$ is taken to be the class of universally measurable functions from X to $\Re^*$.

(b) $\overline{\mathcal{M}}$ is taken to be the class of universally measurable functions from X to U.

(c) g, f, and the probability space of w must satisfy certain Borel measurability conditions.

A key fact is that an ϵ-selection property holds, whereby there exists a $\mu_\epsilon \in \overline{\mathcal{M}}$ such that

$$(T_{\mu_\epsilon} J)(x) \leq \begin{cases} (TJ)(x) + \epsilon & \text{if } (TJ)(x) > -\infty, \\ -(1/\epsilon) & \text{if } (TJ)(x) = -\infty, \end{cases} \qquad \forall\ x \in X, \tag{5.8}$$

† There have been efforts to address the lack of an ϵ-optimal selector within the Borel measurability framework using the concept of a "p-ϵ-optimal selector," whereby the concept of ϵ-optimal selection is modified to hold over a set for states that has p-measure 1, with p being any chosen probability measure over X (see [Str66], [Str75], [DyY79]). This leads to a theory based on p-ϵ-optimal and p-optimal policies, i.e., policies that depend on the choice of p and are optimal only for states in a subset of X that has p-measure 1 (rather than over all states as in our case). It seems difficult to extend the abstract framework of this book based on this inherently probabilistic viewpoint. For a related discussion and a comparison of the p-ϵ-optimal approach with ours, we refer to [BeS78].

for each J in the class $\hat{E}(X)$ of *lower semianalytic* functions from X to $\Re^*$; this is a strict subset of $\overline{E}(X)$, the set of universally measurable functions (see Appendix C).

Because of the difficulty with ϵ-selection within a Borel measurability framework, to construct a more generally applicable restricted policies model, it is necessary to introduce the set of lower semianalytic functions $\hat{E}(X)$ within the Borel space framework, as a *third subset*, additional to $\overline{E}(X)$ and $\overline{\mathcal{M}}$. In summary:

(a) We take $\overline{E}(X)$ to be the class of universally measurable functions from X to $\Re^*$, and $\overline{\mathcal{M}}$ to be the class of universally measurable functions from X to U. Then, $T_\mu J \in \overline{E}(X)$ for all $\mu \in \overline{\mathcal{M}}$ and $J \in \overline{E}(X)$, so the cost function of a policy lies in $\overline{E}(X)$.

(b) We take $\hat{E}(X)$ to be the set of lower semianalytic functions from X to $\Re^*$. Then we have $TJ \in \hat{E}(X)$ for all $J \in \hat{E}(X)$, and the ϵ-selection property (5.8) holds. As a result, the VI algorithm produces functions in $\hat{E}(X)$, if started within in $\hat{E}(X)$, and J^* can be proved to lie in $\hat{E}(X)$.

Motivated by the preceding discussion, we will now introduce a model that involves a set of policies $\overline{\mathcal{M}}$, and two sets of functions $\overline{E}(X)$ and $\hat{E}(X)$ with properties that are analogous to the ones just discussed for Borel space models for stochastic optimal control, along with some additional technical assumptions. We will use this model for the analysis of abstract finite and infinite horizon problems, and we will review later its use within the Borel measurability framework.

5.1.1 General Assumptions

As in Section 4.1, we have the sets X and U, and we introduce a set $\overline{\mathcal{M}}$ of functions $\mu : X \mapsto U$, which we view as a restricted set of stationary policies. We define

$$U(x) = \big\{\mu(x) \mid \mu \in \overline{\mathcal{M}}\big\}, \qquad \forall\, x \in X, \tag{5.9}$$

and we assume that $U(x)$ is nonempty for all $x \in X$. The corresponding set of (nonstationary) policies $\pi = \{\mu_0, \mu_1, \ldots\}$ with $\mu_k \in \overline{\mathcal{M}}$, $k = 0, 1, \ldots$, is denoted by $\overline{\Pi}$. We introduce two subsets $\hat{E}(X)$ and $\overline{E}(X)$ of $E(X)$ (the set of extended real-valued functions on X), such that

$$\hat{E}(X) \subset \overline{E}(X),$$

and the following assumption is satisfied.

Assumption 5.1.1:

(a) For each sequence $\{J_m\} \subset \overline{E}(X)$ with $J_m \to J$, we have $J \in \overline{E}(X)$, and for each sequence $\{J_m\} \subset \hat{E}(X)$ with $J_m \to J$, we have $J \in \hat{E}(X)$.

(b) For all $r \in \Re$, we have

$$J \in \overline{E}(X) \qquad \Rightarrow \qquad J + r\,e \in \overline{E}(X),$$

and

$$J \in \hat{E}(X) \qquad \Rightarrow \qquad J + r\,e \in \hat{E}(X),$$

where e is the unit function, $e(x) \equiv 1$.

We also introduce a mapping $H : X \times U \times \overline{E}(X) \mapsto \Re^*$ satisfying the monotonicity assumption.

Assumption 5.1.2: (Monotonicity) If $J, J' \in \overline{E}(X)$ and $J \leq J'$, then

$$H(x,u,J) \leq H(x,u,J'), \qquad \forall\ x \in X,\ u \in U(x).$$

We define the mapping $T : \overline{E}(X) \mapsto \Re^*$ by

$$(TJ)(x) = \inf_{u \in U(x)} H(x,u,J), \qquad \forall\ x \in X,\ J \in \overline{E}(X),$$

and for each $\mu \in \overline{\mathcal{M}}$, the mapping $T_\mu : \overline{E}(X) \mapsto \Re^*$ by

$$(T_\mu J)(x) = H\big(x, \mu(x), J\big), \qquad \forall\ x \in X,\ J \in \overline{E}(X).$$

Note that in view of the definition (5.9) of $U(x)$, we also have

$$(TJ)(x) = \inf_{\mu \in \overline{\mathcal{M}}} (T_\mu J)(x), \qquad \forall\ x \in X,\ J \in \overline{E}(X),$$

or in shorthand

$$TJ = \inf_{\mu \in \overline{\mathcal{M}}} T_\mu J.$$

The sets $\overline{\mathcal{M}}$, $\overline{E}(X)$, and $\hat{E}(X)$, and the mappings T_μ and T are assumed to satisfy the following.

Assumption 5.1.3:

(a) For all $\mu \in \overline{\mathcal{M}}$, we have

$$J \in \overline{E}(X) \qquad \Rightarrow \qquad T_\mu J \in \overline{E}(X).$$

(b) We have

$$J \in \hat{E}(X) \qquad \Rightarrow \qquad TJ \in \hat{E}(X).$$

Finally, we require that $\hat{E}(X)$ has the following critical ϵ-selection property.

Assumption 5.1.4: (ϵ-Selection) For each $J \in \hat{E}(X)$ and $\epsilon > 0$, there exists a $\mu_\epsilon \in \overline{\mathcal{M}}$ such that

$$(T_{\mu_\epsilon} J)(x) \leq \begin{cases} (TJ)(x) + \epsilon & \text{if } (TJ)(x) > -\infty, \\ -(1/\epsilon) & \text{if } (TJ)(x) = -\infty, \end{cases} \qquad \forall\, x \in X. \quad (5.10)$$

The relevant selection theorem, which guarantees that the Assumption 5.1.4 holds in the stochastic optimal control context is given as Prop. C.5 in Appendix C. Note that Assumption 5.1.3 does not guarantee that $T_\mu J \in \hat{E}(X)$ for $J \in \hat{E}(X)$. As a result, the function $T_{\mu_\epsilon} J$ of Assumption 5.1.4 is only guaranteed to belong to the larger set $\overline{E}(X)$.

Problem Formulation

We are given a function $\bar{J} \in \hat{E}(X)$ satisfying

$$\bar{J}(x) > -\infty, \qquad \forall\, x \in X, \quad (5.11)$$

and we consider for every policy $\pi = \{\mu_0, \mu_1, \ldots\} \in \overline{\Pi}$ and positive integer N the function $J_{N,\pi} \in \overline{E}(X)$ defined by

$$J_{N,\pi}(x) = (T_{\mu_0} \cdots T_{\mu_{N-1}} \bar{J})(x), \qquad \forall\, x \in X,$$

and the function J_π defined by

$$J_\pi(x) = \limsup_{k \to \infty} (T_{\mu_0} \cdots T_{\mu_k} \bar{J})(x), \qquad \forall\, x \in X.$$

For a stationary policy $\pi = \{\mu, \mu, \ldots\}$ we also write J_μ in place of J_π.

As earlier, we consider the N-stage optimization problem

$$\begin{array}{ll} \text{minimize} & J_{N,\pi}(x) \\ \text{subject to} & \pi \in \overline{\Pi}, \end{array} \tag{5.12}$$

and its infinite horizon version

$$\begin{array}{ll} \text{minimize} & J_{\pi}(x) \\ \text{subject to} & \pi \in \overline{\Pi}. \end{array} \tag{5.13}$$

For a fixed $x \in X$, we denote by $J_N^*(x)$ and $J^*(x)$ the optimal costs for these problems, i.e.,

$$J_N^*(x) = \inf_{\pi \in \overline{\Pi}} J_{N,\pi}(x), \qquad J^*(x) = \inf_{\pi \in \overline{\Pi}} J_{\pi}(x), \qquad \forall\ x \in X.$$

We say that a policy $\pi^* \in \overline{\Pi}$ is *N-stage optimal* if

$$J_{N,\pi^*}(x) = J_N^*(x), \qquad \forall\ x \in X,$$

and (infinite horizon) *optimal* if

$$J_{\pi^*}(x) = J^*(x), \qquad \forall\ x \in X.$$

For a given $\epsilon > 0$, we say that $\pi_\epsilon \in \overline{\Pi}$ is *N-stage ϵ-optimal* if

$$J_{\pi_\epsilon}(x) \leq \begin{cases} J_N^*(x) + \epsilon & \text{if } J_N^*(x) > -\infty, \\ -(1/\epsilon) & \text{if } J_N^*(x) = -\infty, \end{cases}$$

and we say that π_ϵ is (infinite horizon) *ϵ-optimal* if

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

Note that since $\bar{J} \in \hat{E}(X)$, the function $T^k \bar{J}$ belongs to $\hat{E}(X)$ for all k [cf. Assumption 5.1.3(b)]. Similar to Chapter 4, we will aim to show under various assumptions that $J_N^* = T^N \bar{J}$, and that $J^* \in \hat{E}(X)$ and $J^* = TJ^*$.

5.2 FINITE HORIZON PROBLEMS

To show that $J_N^* = T^N \bar{J}$, we use an analysis that is similar to the one of Section 4.2. In particular, we introduce the following assumption, which is analogous to Assumption 4.2.1.

Assumption 5.2.1: For each sequence $\{J_m\} \subset \overline{E}(X)$ with $J_m \downarrow J$ for some $J \in \overline{E}(X)$, we have

$$\lim_{m\to\infty} H(x,u,J_m) \leq H(x,u,J), \qquad \forall\ x \in X,\ u \in U(x). \tag{5.14}$$

Note that Assumption 5.1.1 implies that if $\{J_m\} \subset \overline{E}(X)$ and $J_m \downarrow J \in \overline{E}(X)$, we have

$$J = \lim_{m\to\infty} J_m = \inf_{m=0,1,\ldots} J_m \in \overline{E}(X),$$

so for all $\mu \in \overline{\mathcal{M}}$, by Assumption 5.2.1,

$$\inf_m (T_\mu J_m) = T_\mu \left(\inf_m J_m \right).$$

This inequality can be extended for any $\mu_1, \ldots, \mu_k \in \overline{\mathcal{M}}$ as follows:

$$\begin{aligned} \inf_m (T_{\mu_1} \cdots T_{\mu_k} J_m) &= T_{\mu_1} \left(\inf_m (T_{\mu_1} \cdots T_{\mu_k} J_m) \right) \\ &= \cdots \\ &= T_{\mu_1} \cdots T_{\mu_k} \left(\inf_m J_m \right). \end{aligned} \tag{5.15}$$

We have the following proposition, which extends Prop. 4.2.3 to the restricted policies framework.

Proposition 5.2.1: Let Assumptions 5.1.1-5.1.4 and 5.2.1 hold. Then

$$J_N^* = T^N \bar{J}.$$

Proof: We select for each $k = 0, \ldots, N-1$, a sequence $\{\mu_k^m\} \subset \overline{\mathcal{M}}$ such that

$$\lim_{m\to\infty} T_{\mu_k^m} (T^{N-k-1} \bar{J}) \downarrow T^{N-k} \bar{J}.$$

This is possible in view of the ϵ-selection property of Assumption 5.1.4, since $T^{N-k}\bar{J} \in \hat{E}(X)$ [cf. Assumption 5.1.3(b)]. Using Eq. (5.15), we have

$$\begin{aligned} J_N^* &\leq \inf_{m_0} \cdots \inf_{m_{N-1}} T_{\mu_0^{m_0}} \cdots T_{\mu_{N-1}^{m_{N-1}}} \bar{J} \\ &= \inf_{m_0} \cdots \inf_{m_{N-2}} T_{\mu_0^{m_0}} \cdots T_{\mu_{N-2}^{m_{N-2}}} \left(\inf_{m_{N-1}} T_{\mu_{N-1}^{m_{N-1}}} \bar{J} \right) \\ &= \inf_{m_0} \cdots \inf_{m_{N-2}} T_{\mu_0^{m_0}} \cdots T_{\mu_{N-2}^{m_{N-2}}} T\bar{J} \\ &= \cdots \\ &= T^N \bar{J}, \end{aligned}$$

where the last equality is obtained by repeating the process used to obtain the previous equalities. On the other hand, it is clear from the definitions that $T^N\bar{J} \leq J_{N,\pi}$ for all N and $\pi \in \Pi$, so that $T^N\bar{J} \leq J_N^*$. Thus, $J_N^* = T^N\bar{J}$. **Q.E.D.**

We also introduce the following alternative assumption, which parallels Assumption 4.2.2.

Assumption 5.2.2: The k-stages optimal cost function J_k^* satisfies

$$J_k^*(x) > -\infty, \qquad \forall\, x \in X,\ k = 1, \ldots, N.$$

Moreover, there exists a scalar $\alpha \in (0, \infty)$ such that for all scalars $r \in (0, \infty)$ and functions $J \in \overline{E}(X)$, we have

$$H(x, u, J) \leq H(x, u, J + r\, e) \leq H(x, u, J) + \alpha\, r, \qquad \forall\, x \in X,\ u \in U(x). \tag{5.16}$$

We have the following proposition whose statement and proof parallel the ones of Prop. 4.2.4.

Proposition 5.2.2: Let Assumptions 5.1.1-5.1.4, and 5.2.2 hold. Then $J_N^* = T^N\bar{J}$, and for every $\epsilon > 0$, there exists an ϵ-optimal policy.

Proof: Note that since by assumption, $J_N^*(x) > -\infty$ for all $x \in X$, an N-stage ϵ-optimal policy $\pi_\epsilon \in \overline{\Pi}$ is one for which

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

We use induction. The result clearly holds for $N = 1$. Assume that it holds for $N = k$, i.e., $J_k^* = T^k\bar{J} \in \hat{E}(X)$ and for a given $\epsilon > 0$, there is a

$\pi_e \in \overline{\Pi}$ with $J_{k,\pi_\epsilon} \leq J_k^* + \epsilon\, e$. Using Eq. (5.16), we have for all $\mu \in \overline{\mathcal{M}}$,

$$J_{k+1}^* \leq T_\mu J_{k,\pi_\epsilon} \leq T_\mu J_k^* + \alpha\epsilon\, e.$$

Taking the infimum over μ and then the limit as $\epsilon \to 0$, we obtain $J_{k+1}^* \leq TJ_k^*$. By using the induction hypothesis, it follows that $J_{k+1}^* \leq T^{k+1}\bar{J}$. On the other hand, we have clearly $T^{k+1}\bar{J} \leq J_{k+1}^*$, and hence $T^{k+1}\bar{J} = J_{k+1}^*$.

Using the assumption $J_k^*(x) > -\infty$ for all $x \in X$, for any $\overline{\epsilon} > 0$, we can choose $\overline{\pi} = \{\overline{\mu}_0, \overline{\mu}_1, \ldots\} \in \overline{\Pi}$ such that

$$J_{k,\pi} \leq J_k^* + \frac{\overline{\epsilon}}{2\alpha}\, e,$$

and $\overline{\mu} \in \overline{\mathcal{M}}$ such that

$$T_{\overline{\mu}} J_k^* \leq TJ_k^* + \frac{\overline{\epsilon}}{2}\, e.$$

Let $\overline{\pi}_{\overline{\epsilon}} = \{\overline{\mu}, \overline{\mu}_0, \overline{\mu}_1, \ldots\}$. Then

$$J_{k+1,\overline{\pi}_{\overline{\epsilon}}} = T_{\overline{\mu}} J_{k,\overline{\pi}} \leq T_{\overline{\mu}} J_k^* + \frac{\overline{\epsilon}}{2}\, e \leq TJ_k^* + \overline{\epsilon}\, e = J_{k+1}^* + \overline{\epsilon}\, e,$$

where the first inequality is obtained by using Eq. (5.16). The induction is complete. **Q.E.D.**

5.3 CONTRACTIVE MODELS

In this section, we will discuss briefly the infinite horizon problem

$$\begin{aligned} &\text{minimize} \quad J_\pi(x) \\ &\text{subject to} \quad \pi \in \overline{\Pi}, \end{aligned} \tag{5.17}$$

where for a policy $\{\mu_0, \mu_1, \ldots\} \in \overline{\Pi}$, $J_\pi \in \overline{E}(X)$ is defined by

$$J_\pi(x) = \limsup_{k\to\infty} (T_{\mu_0} \cdots T_{\mu_k} \bar{J})(x), \qquad \forall\ x \in X.$$

We analyze this problem under a contraction assumption, similar to Chapter 2. To this end, we introduce a function $v : X \mapsto \Re$ with

$$v(x) > 0, \qquad \forall\ x \in X,$$

and we consider the weighted sup-norm

$$\|J\| = \sup_{x\in X} \frac{|J(x)|}{v(x)}$$

on $B(X)$, the space of real-valued functions J on X such that $J(x)/v(x)$ is bounded over $x \in X$.

In addition to the Assumptions 5.1.1-5.1.4, we assume the following.

Assumption 5.3.1: (Contraction)

(a) The sets $\overline{E}(X)$ and $\hat{E}(X)$ are closed subsets of $B(X)$.

(b) For some $\alpha \in (0,1)$, we have

$$\|T_\mu J - T_\mu J'\| \leq \alpha \|J - J'\|, \qquad \forall\ J, J' \in \overline{E}(X),\ \mu \in \overline{\mathcal{M}}. \quad (5.18)$$

The analysis of Chapter 2 carries through with few modifications. In particular the following analog of the major analytical result of Chapter 2, can be proved with an essentially identical proof. The key fact here is that $\overline{E}(X)$ and $\hat{E}(X)$ are closed subsets of $B(X)$, so the Contraction Mapping Theorem (Prop. B.1) applies.

Proposition 5.3.1: Let Assumptions 5.1.1-5.1.4 and the contraction Assumption 5.3.1 hold. Then:

(a) For all $\mu \in \overline{\mathcal{M}}$, the mapping T_μ is a contraction mapping with modulus α over $\overline{E}(X)$, and its unique fixed point within $\overline{E}(X)$ is J_μ.

(b) The mapping T is a contraction mapping with modulus α over $\hat{E}(X)$, and its unique fixed point within $\hat{E}(X)$ is equal to J^*.

(c) For any $J \in \overline{E}(X)$ and $\mu \in \overline{\mathcal{M}}$, we have

$$\lim_{k\to\infty} T_\mu^k J = J_\mu.$$

(d) For any $J \in \hat{E}(X)$, we have

$$\lim_{k\to\infty} T^k J = J^*.$$

(e) We have $T_\mu J^* = TJ^*$ if and only if $J_\mu = J^*$.

(f) For every $\epsilon > 0$ there exists an ϵ-optimal policy within $\overline{\mathcal{M}}$.

Proof: As in Section 1.2, T_μ is a contraction with modulus α over $\overline{E}(X)$. Similarly, T is a contraction with modulus α over $\hat{E}(X)$. Parts (a), (b),

(c), and (d) follow from Prop. B.1 of Appendix B.

To show part (e), note that if $T_\mu J^* = TJ^*$, then in view of $TJ^* = J^*$, we have $T_\mu J^* = J^*$, which implies that $J^* = J_\mu$, since J_μ is the unique fixed point of T_μ. Conversely, if $J_\mu = J^*$, we have

$$T_\mu J^* = T_\mu J_\mu = J_\mu = J^* = TJ^*.$$

Part (f) follows similarly, using the proof of Prop. 2.1.2. **Q.E.D.**

5.4 BOREL SPACE MODELS

We will now apply the preceding analysis to models where the set of policies is restricted in order to address measurability issues. The Borel space model is the most general such model, and we will focus on it. Appendix C provides a motivation and an outline of the model for finite horizon problems, including the associated mathematical definitions, some basic results, and a two-stage example. In this section we will provide a brief discussion of an infinite horizon contractive model.

We consider the mapping H defined by

$$H(x,u,J) = g(x,u) + \alpha \int J(y)\, p(dy \mid x,u). \tag{5.19}$$

Here X and U are Borel spaces, α is a scalar in $(0,1]$, J is an extended real-valued functions on X, and $p(dy \mid x,u)$ is a transition probability measure for each x and $u \in U(x)$. To make mathematical sense of the expression in the right-hand side of Eq. (5.19), J must satisfy certain measurability restrictions, so we assume that g is Borel measurable and that $p(dy \mid x,u)$ is a Borel measurable stochastic kernel. We let

$\overline{E}(X) =$ the subset of universally measurable functions from $E(X)$.

Then the integral in Eq. (5.19) is well-defined as an extended real number for every (x,u) for all $J \in \overline{E}(X)$ (recall that in our integration framework we allow the sum $\infty - \infty$ to appear and interpret it as ∞).

A requirement of the framework of this chapter is that $T_\mu J$ must belong to $\overline{E}(X)$ for each $J \in \overline{E}(X)$. For this it is sufficient that g be Borel measurable as a function of (x,u), and the policy μ be a universally measurable function of x. However, as noted earlier, universal measurability of H [as a function of (x,u) for fixed J] is insufficient to guarantee that $TJ \in \overline{E}(X)$. We thus let

$\hat{E}(X) =$ the subset of lower semianalytic functions from $\overline{E}(X)$.

Then assuming that $J \in \hat{E}(X)$, that $g(x,u)$ and $p(dy \mid x,u)$ are Borel measurable, and that the set

$$\big\{(x,u) \mid u \in U(x)\big\}$$

is Borel measurable, we have that the function $H(\cdot,\cdot,J)$ is lower semianalytic and $TJ \in \hat{E}(X)$, as discussed in Appendix C (cf. Prop. C.4).

Note that our framework requires that ϵ-optimal selection is possible, i.e., that for every $\epsilon > 0$, there exists a universally measurable μ such that the conditions of Assumption 5.1.4 are satisfied. This is ensured under our assumptions by the selection theorem of Prop. C.5 in Appendix C.

We now consider a contractive infinite horizon problem, which is based on the mapping H defined by Eq. (5.19), where $\alpha \in (0,1)$ is a discount factor, and the preceding conditions hold. In addition g is assumed to be not only Borel measurable, but also bounded above and below, in which case we obtain a contractive model in the space of bounded functions $B(X)$ with respect to the unweighted sup-norm. It is important to note that $\overline{E}(X)$ and $\hat{E}(X)$ are closed subsets of $B(X)$, since the pointwise limit of universally measurable and lower semianalytic functions are universally measurable and lower semianalytic, respectively (see [BeS78]).

Thus Prop. 5.3.1 applies and provides the basic analytical results for contractive Borel space models, which are:

(a) J^* is the unique fixed point of T within $\hat{E}(X)$.

(b) For every $J \in \hat{E}(X)$, $T^k J \in \hat{E}(X)$ for all k, and we have $T^k J \to J^*$.

(c) There exists a universally measurable optimal policy if and only if the infimum of $H(x,u,J^*)$ over u is attained for each $x \in X$.

(d) For any $\epsilon > 0$, there exists an ϵ-optimal universally measurable policy.

For a detailed discussion and proofs of these results, we refer to [BeS78].

5.5 NOTES, SOURCES, AND EXERCISES

The restricted model framework of this chapter was treated briefly in the book [BeS78] (Chapter 6). This book focused far more extensively on the classical type of stochastic optimal control problems (cf. Example 1.2.1), rather than the more general abstract restricted model case.

The restricted policies framework may also be applied to the so-called *semicontinuous models* similar to how it was applied to Borel models in Section 5.4. The semicontinuous models provide more powerful results regarding the character of the cost functions, but require additional assumptions, which may be restrictive, namely that the cost function g and the stochastic kernel p in Eq. (5.19) have certain upper or lower semicontinuity properties. The relevant mathematical background is given in Section 7.5 of [BeS78],

and the critical selection theorems (with Borel measurable selection) are given in Props. 7.33 and 7.34 of that reference. Detailed related references to the literature may also be found in [BeS78].