APPENDIX A:

Notation and Mathematical Conventions

In this appendix we collect our notation, and some related mathematical facts and conventions.

A.1 SET NOTATION AND CONVENTIONS

If X is a set and x is an element of X, we write $x \in X$. A set can be specified in the form $X = \{x \mid x \text{ satisfies } P\}$, as the set of all elements satisfying property P. The union of two sets X_1 and X_2 is denoted by $X_1 \cup X_2$, and their intersection by $X_1 \cap X_2$. The empty set is denoted by $\emptyset$. The symbol $\forall$ means "for all."

The set of real numbers (also referred to as scalars) is denoted by $\Re$. The set of extended real numbers is denoted by $\Re^*$:

$$\Re^* = \Re \cup \{\infty, -\infty\}.$$

We write $-\infty < x < \infty$ for all real numbers x, and $-\infty \leq x \leq \infty$ for all extended real numbers x. We denote by $[a, b]$ the set of (possibly extended) real numbers x satisfying $a \leq x \leq b$. A rounded, instead of square, bracket denotes strict inequality in the definition. Thus $(a, b]$, $[a, b)$, and (a, b) denote the set of all x satisfying $a < x \leq b$, $a \leq x < b$, and $a < x < b$, respectively.

Generally, we adopt standard conventions regarding addition and multiplication in $\Re^*$, except that we take

$$\infty - \infty = -\infty + \infty = \infty,$$

and we take the product of 0 and ∞ or $-\infty$ to be 0. In this way the sum and product of two extended real numbers is well-defined. Division by 0 or ∞ does not appear in our analysis. In particular, we adopt the following rules in calculations involving ∞ and $-\infty$:

$$\alpha + \infty = \infty + \alpha = \infty, \qquad \forall\ \alpha \in \Re^*,$$

$$\alpha - \infty = -\infty + \alpha = -\infty, \qquad \forall\ \alpha \in [-\infty, \infty),$$

$$\alpha \cdot \infty = \infty, \qquad \alpha \cdot (-\infty) = \infty, \qquad \forall\ \alpha \in (0, \infty],$$

$$\alpha \cdot \infty = -\infty, \qquad \alpha \cdot (-\infty) = -\infty, \qquad \forall\ \alpha \in [-\infty, 0),$$

$$0 \cdot \infty = \infty \cdot 0 = 0 = 0 \cdot (-\infty) = (-\infty) \cdot 0, \qquad -(-\infty) = \infty.$$

Under these rules, the following laws of arithmetic are still valid within $\Re^*$:

$$\alpha_1 + \alpha_2 = \alpha_2 + \alpha_1, \qquad (\alpha_1 + \alpha_2) + \alpha_3 = \alpha_1 + (\alpha_2 + \alpha_3),$$

$$\alpha_1 \alpha_2 = \alpha_2 \alpha_1, \qquad (\alpha_1 \alpha_2)\alpha_3 = \alpha_1(\alpha_2 \alpha_3).$$

We also have

$$\alpha(\alpha_1 + \alpha_2) = \alpha\alpha_1 + \alpha\alpha_2$$

if either $\alpha \geq 0$ or else $(\alpha_1 + \alpha_2)$ is not of the form $\infty - \infty$.

Inf and Sup Notation

The *supremum* of a nonempty set $X \subset \Re^*$, denoted by $\sup X$, is defined as the smallest $y \in \Re^*$ such that $y \geq x$ for all $x \in X$. Similarly, the *infimum* of X, denoted by $\inf X$, is defined as the largest $y \in \Re^*$ such that $y \leq x$ for all $x \in X$. For the empty set, we use the convention

$$\sup \emptyset = -\infty, \qquad \inf \emptyset = \infty.$$

If $\sup X$ is equal to an $\overline{x} \in \Re^*$ that belongs to the set X, we say that $\overline{x}$ is the *maximum point* of X and we write $\overline{x} = \max X$. Similarly, if $\inf X$ is equal to an $\overline{x} \in \Re^*$ that belongs to the set X, we say that $\overline{x}$ is the *minimum point* of X and we write $\overline{x} = \min X$. Thus, when we write $\max X$ (or $\min X$) in place of $\sup X$ (or $\inf X$, respectively), we do so just for emphasis: we indicate that it is either evident, or it is known through earlier analysis, or it is about to be shown that the maximum (or minimum, respectively) of the set X is attained at one of its points.

A.2 FUNCTIONS

If f is a function, we use the notation $f : X \mapsto Y$ to indicate the fact that f is defined on a nonempty set X (its *domain*) and takes values in a set Y (its *range*). Thus when using the notation $f : X \mapsto Y$, we implicitly assume that X is nonempty. We will often use the *unit function* $e : X \mapsto \Re$, defined by

$$e(x) = 1, \qquad \forall\ x \in X.$$

Given a set X, we denote by $R(X)$ the set of real-valued functions $J : X \mapsto \Re$, and by $E(X)$ the set of all extended real-valued functions $J : X \mapsto \Re^*$. For any collection $\{J_\gamma \mid \gamma \in \Gamma\} \subset E(X)$, parameterized by the elements of a set Γ, we denote by $\inf_{\gamma\in\Gamma} J_\gamma$ the function taking the value $\inf_{\gamma\in\Gamma} J_\gamma(x)$ at each $x \in X$.

For two functions $J_1, J_2 \in E(X)$, we use the shorthand notation $J_1 \le J_2$ to indicate the pointwise inequality

$$J_1(x) \le J_2(x), \qquad \forall\ x \in X.$$

We use the shorthand notation $\inf_{i\in I} J_i$ to denote the function obtained by pointwise infimum of a collection $\{J_i \mid i \in I\} \subset E(X)$, i.e.,

$$\left(\inf_{i\in I} J_i\right)(x) = \inf_{i\in I} J_i(x), \qquad \forall\ x \in X.$$

We use similar notation for sup.

Given subsets $S_1, S_2, S_3 \subset E(X)$ and mappings $T_1 : S_1 \mapsto S_3$ and $T_2 : S_2 \mapsto S_1$, the *composition* of T_1 and T_2 is the mapping $T_1T_2 : S_2 \mapsto S_3$ defined by

$$(T_1T_2J)(x) = \big(T_1(T_2J)\big)(x), \qquad \forall\ J \in S_2,\ x \in X.$$

In particular, given a subset $S \subset E(X)$ and mappings $T_1 : S \mapsto S$ and $T_2 : S \mapsto S$, the composition of T_1 and T_2 is the mapping $T_1T_2 : S \mapsto S$ defined by

$$(T_1T_2J)(x) = \big(T_1(T_2J)\big)(x), \qquad \forall\ J \in S,\ x \in X.$$

Similarly, given mappings $T_k : S \mapsto S$, $k = 1, \ldots, N$, their composition is the mapping $(T_1 \cdots T_N) : S \mapsto S$ defined by

$$(T_1T_2 \cdots T_NJ)(x) = \big(T_1(T_2(\cdots(T_NJ)))\big)(x), \qquad \forall\ J \in S,\ x \in X.$$

In our notation involving compositions we minimize the use of parentheses, as long as there is no ambiguity. Thus we write T_1T_2J instead of (T_1T_2J) or $(T_1T_2)J$ or $T_1(T_2J)$, but we write $(T_1T_2J)(x)$ to indicate the value of T_1T_2J at $x \in X$.

If X and Y are nonempty sets, a mapping $T : S_1 \mapsto S_2$, where $S_1 \subset E(X)$ and $S_2 \subset E(Y)$, is said to be *monotone* if for all $J, J' \in S_1$,

$$J \le J' \qquad \Rightarrow \qquad TJ \le TJ'.$$

Sequences of Functions

For a sequence of functions $\{J_k\} \subset E(X)$ that converges pointwise, we denote by $\lim_{k\to\infty} J_k$ the pointwise limit of $\{J_k\}$. We denote by $\limsup_{k\to\infty} J_k$ (or $\liminf_{k\to\infty} J_k$) the pointwise limit superior (or inferior, respectively) of $\{J_k\}$. If $\{J_k\} \subset E(X)$ converges pointwise to J, we write $J_k \to J$. Note that we reserve this notation for pointwise convergence. To denote convergence with respect to a norm $\|\cdot\|$, we write $\|J_k - J\| \to 0$.

A sequence of functions $\{J_k\} \subset E(X)$ is said to be *monotonically nonincreasing* (or *monotonically nondecreasing*) if $J_{k+1} \leq J_k$ for all k (or $J_{k+1} \geq J_k$ for all k, respectively). Such a sequence always has a (pointwise) limit within $E(X)$. We write $J_k \downarrow J$ (or $J_k \uparrow J$) to indicate that $\{J_k\}$ is monotonically nonincreasing (or monotonically nonincreasing, respectively) and that its limit is J.

Let $\{J_{mn}\} \subset E(X)$ be a double indexed sequence, which is monotonically nonincreasing separately for each index in the sense that

$$J_{(m+1)n} \leq J_{mn}, \qquad J_{m(n+1)} \leq J_{mn}, \qquad \forall\ m, n = 0, 1, \ldots.$$

For such sequences, a useful fact is that

$$\lim_{m\to\infty} \left(\lim_{n\to\infty} J_{mn} \right) = \lim_{m\to\infty} J_{mm}.$$

There is a similar fact for monotonically nondecreasing sequences.

Expected Values

Given a random variable w defined over a probability space Ω, the expected value of w is defined by

$$E\{w\} = E\{w^+\} + E\{w^-\},$$

where w^+ and w^- are the positive and negative parts of w,

$$w^+(\omega) = \max\big\{0, w(\omega)\big\}, \qquad w^-(\omega) = \min\big\{0, w(\omega)\big\}.$$

In this way, taking also into account the rule $\infty - \infty = \infty$, the expected value $E\{w\}$ is well-defined if Ω is finite or countably infinite. In more general cases, $E\{w\}$ is similarly defined by the appropriate form of integration, as will be discussed in more detail at specific points as needed.

APPENDIX B:

Contraction Mappings

B.1 CONTRACTION MAPPING FIXED POINT THEOREMS

The purpose of this appendix is to provide some background on contraction mappings and their properties. Let Y be a real vector space with a norm $\|\cdot\|$, i.e., a real-valued function satisfying for all $y \in Y$, $\|y\| \geq 0$, $\|y\| = 0$ if and only if $y = 0$, and

$$\|ay\| = |a|\|y\|, \quad \forall\ a \in \Re, \qquad \|y+z\| \leq \|y\| + \|z\|, \quad \forall\ y, z \in Y.$$

Let $\bar{Y}$ be a closed subset of Y. A function $F : \bar{Y} \mapsto \bar{Y}$ is said to be a *contraction mapping* if for some $\rho \in (0,1)$, we have

$$\|Fy - Fz\| \leq \rho\|y - z\|, \qquad \forall\ y, z \in \bar{Y}.$$

The scalar ρ is called the *modulus of contraction* of F.

Example B.1 (Linear Contraction Mappings in $\Re^n$)

Consider the case of a linear mapping $F : \Re^n \mapsto \Re^n$ of the form

$$Fy = b + Ay,$$

where A is an $n \times n$ matrix and b is a vector in $\Re^n$. Let $\sigma(A)$ denote the spectral radius of A (the largest modulus among the moduli of the eigenvalues of A). Then it can be shown that A *is a contraction mapping with respect to some norm if and only if* $\sigma(A) < 1$.

Specifically, given $\epsilon > 0$, there exists a norm $\|\cdot\|_s$ such that

$$\|Ay\|_s \leq \big(\sigma(A) + \epsilon\big)\|y\|_s, \qquad \forall\ y \in \Re^n. \tag{B.1}$$

Thus, if $\sigma(A) < 1$ we may select $\epsilon > 0$ such that $\rho = \sigma(A) + \epsilon < 1$, and obtain the contraction relation

$$\|Fy - Fz\|_s = \big\|A(y - z)\big\|_s \leq \rho\|y - z\|_s, \qquad \forall\ y, z \in \Re^n. \tag{B.2}$$

The norm $\|\cdot\|_s$ can be taken to be a weighted Euclidean norm, i.e., it may have the form $\|y\|_s = \|My\|$, where M is a square invertible matrix, and $\|\cdot\|$ is the standard Euclidean norm, i.e., $\|x\| = \sqrt{x'x}$. †

Conversely, if Eq. (B.2) holds for some norm $\|\cdot\|_s$ and all real vectors y, z, it also holds for all complex vectors y, z with the squared norm $\|c\|_s^2$ of a complex vector c defined as the sum of the squares of the norms of the real and the imaginary components. Thus from Eq. (B.2), by taking $y - z = u$, where u is an eigenvector corresponding to an eigenvalue λ with $|\lambda| = \sigma(A)$, we have $\sigma(A)\|u\|_s = \|Au\|_s \leq \rho\|u\|_s$. Hence $\sigma(A) \leq \rho$, and it follows that if F is a contraction with respect to a given norm, we must have $\sigma(A) < 1$.

A sequence $\{y_k\} \subset Y$ is said to be a *Cauchy sequence* if $\|y_m - y_n\| \to 0$ as $m, n \to \infty$, i.e., given any $\epsilon > 0$, there exists N such that $\|y_m - y_n\| \leq \epsilon$ for all $m, n \geq N$. The space Y is said to be *complete* under the norm $\|\cdot\|$ if every Cauchy sequence $\{y_k\} \subset Y$ is convergent, in the sense that for some $\bar{y} \in Y$, we have $\|y_k - \bar{y}\| \to 0$. Note that a Cauchy sequence is always bounded. Also, a Cauchy sequence of real numbers is convergent, implying that the real line is a complete space and so is every real finite-dimensional vector space. On the other hand, an infinite dimensional space may not be complete under some norms, while it may be complete under other norms.

When Y is complete and $\bar{Y}$ is a closed subset of Y, an important property of a contraction mapping $F : \bar{Y} \mapsto \bar{Y}$ is that it has a unique fixed point within $\bar{Y}$, i.e., the equation

$$y = Fy$$

has a unique solution $y^* \in \bar{Y}$, called the *fixed point of* F. Furthermore, the sequence $\{y_k\}$ generated by the iteration

$$y_{k+1} = Fy_k$$

---

† We may show Eq. (B.1) by using the Jordan canonical form of A, which is denoted by J. In particular, if P is a nonsingular matrix such that $P^{-1}AP = J$ and D is the diagonal matrix with $1, \delta, \ldots, \delta^{n-1}$ along the diagonal, where $\delta > 0$, it is straightforward to verify that $D^{-1}P^{-1}APD = \hat{J}$, where $\hat{J}$ is the matrix that is identical to J except that each nonzero off-diagonal term is replaced by δ. Defining $\hat{P} = PD$, we have $A = \hat{P}\hat{J}\hat{P}^{-1}$. Now if $\|\cdot\|$ is the standard Euclidean norm, we note that for some $\beta > 0$, we have $\|\hat{J}z\| \leq \big(\sigma(A) + \beta\delta\big)\|z\|$ for all $z \in \Re^n$ and $\delta \in (0, 1]$. For a given $\delta \in (0, 1]$, consider the weighted Euclidean norm $\|\cdot\|_s$ defined by $\|y\|_s = \|\hat{P}^{-1}y\|$. Then we have for all $y \in \Re^n$,

$$\|Ay\|_s = \|\hat{P}^{-1}Ay\| = \|\hat{P}^{-1}\hat{P}\hat{J}\hat{P}^{-1}y\| = \|\hat{J}\hat{P}^{-1}y\| \leq \big(\sigma(A) + \beta\delta\big)\|\hat{P}^{-1}y\|,$$

so that $\|Ay\|_s \leq \big(\sigma(A) + \beta\delta\big)\|y\|_s$, for all $y \in \Re^n$. For a given $\epsilon > 0$, we choose $\delta = \epsilon/\beta$, so the preceding relation yields Eq. (B.1).

converges to y^*, starting from an arbitrary initial point y_0.

Proposition B.1: (Contraction Mapping Fixed-Point Theorem) Let Y be a complete vector space and let $\bar{Y}$ be a closed subset of Y. Then if $F : \bar{Y} \mapsto \bar{Y}$ is a contraction mapping with modulus $\rho \in (0,1)$, there exists a unique $y^* \in \bar{Y}$ such that

$$y^* = Fy^*.$$

Furthermore, the sequence $\{F^k y\}$ converges to y^* for any $y \in \bar{Y}$, and we have

$$\|F^k y - y^*\| \le \rho^k \|y - y^*\|, \qquad k = 1, 2, \ldots.$$

Proof: Let $y \in \bar{Y}$ and consider the iteration $y_{k+1} = Fy_k$ starting with $y_0 = y$. By the contraction property of F,

$$\|y_{k+1} - y_k\| \le \rho \|y_k - y_{k-1}\|, \qquad k = 1, 2, \ldots,$$

which implies that

$$\|y_{k+1} - y_k\| \le \rho^k \|y_1 - y_0\|, \qquad k = 1, 2, \ldots.$$

It follows that for every $k \ge 0$ and $m \ge 1$, we have

$$\begin{aligned} \|y_{k+m} - y_k\| &\le \sum_{i=1}^{m} \|y_{k+i} - y_{k+i-1}\| \\ &\le \rho^k (1 + \rho + \cdots + \rho^{m-1}) \|y_1 - y_0\| \\ &\le \frac{\rho^k}{1-\rho} \|y_1 - y_0\|. \end{aligned}$$

Therefore, $\{y_k\}$ is a Cauchy sequence in $\bar{Y}$ and must converge to a limit $y^* \in \bar{Y}$, since Y is complete and $\bar{Y}$ is closed. We have for all $k \ge 1$,

$$\|Fy^* - y^*\| \le \|Fy^* - y_k\| + \|y_k - y^*\| \le \rho \|y^* - y_{k-1}\| + \|y_k - y^*\|$$

and since y_k converges to y^*, we obtain $Fy^* = y^*$. Thus, the limit y^* of y_k is a fixed point of F. It is a unique fixed point because if $\tilde{y}$ were another fixed point, we would have

$$\|y^* - \tilde{y}\| = \|Fy^* - F\tilde{y}\| \le \rho \|y^* - \tilde{y}\|,$$

which implies that $y^* = \tilde{y}$.

To show the convergence rate bound of the last part, note that

$$\|F^k y - y^*\| = \big\|F^k y - F y^*\big\| \leq \rho \|F^{k-1} y - y^*\|.$$

Repeating this process for a total of k times, we obtain the desired result. **Q.E.D.**

The convergence rate exhibited by $F^k y$ in the preceding proposition is said to be *geometric*, and $F^k y$ is said to converge to its limit y^* *geometrically*. This is in reference to the fact that the error $\|F^k y - y^*\|$ converges to 0 faster than some geometric progression ($\rho^k \|y - y^*\|$ in this case).

In some contexts of interest to us one may encounter mappings that are not contractions, but become contractions when iterated a finite number of times. In this case, one may use a slightly different version of the contraction mapping fixed point theorem, which we now present.

We say that a function $F : \bar{Y} \mapsto \bar{Y}$ is an *m-stage contraction mapping* if there exists a positive integer m and some $\rho < 1$ such that

$$\big\|F^m y - F^m y'\big\| \leq \rho \|y - y'\|, \qquad \forall\, y, y' \in \bar{Y},$$

where F^m denotes the composition of F with itself m times. Thus, F is an m-stage contraction if F^m is a contraction. Again, the scalar ρ is called the modulus of contraction. We have the following generalization of Prop. B.1.

Proposition B.2: (m-Stage Contraction Mapping Fixed-Point Theorem) Let Y be a complete vector space and let $\bar{Y}$ be a closed subset of Y. Then if $F : \bar{Y} \mapsto \bar{Y}$ is an m-stage contraction mapping with modulus $\rho \in (0, 1)$, there exists a unique $y^* \in \bar{Y}$ such that

$$y^* = F y^*.$$

Furthermore, $\{F^k y\}$ converges to y^* for any $y \in \bar{Y}$.

Proof: Since F^m maps $\bar{Y}$ into $\bar{Y}$ and is a contraction mapping, by Prop. B.1, it has a unique fixed point in $\bar{Y}$, denoted y^*. Applying F to both sides of the relation $y^* = F^m y^*$, we see that $F y^*$ is also a fixed point of F^m, so by the uniqueness of the fixed point, we have $y^* = F y^*$. Therefore y^* is a fixed point of F. If F had another fixed point, say $\tilde{y}$, then we would have $\tilde{y} = F^m \tilde{y}$, which by the uniqueness of the fixed point of F^m implies that $\tilde{y} = y^*$. Thus, y^* is the unique fixed point of F.

To show the convergence of $\{F^k y\}$, note that by Prop. B.1, we have for all $y \in \bar{Y}$,

$$\lim_{k \to \infty} \|F^{mk} y - y^*\| = 0.$$

Using $F^\ell y$ in place of y, we obtain

$$\lim_{k\to\infty} \|F^{mk+\ell}y - y^*\| = 0, \qquad \ell = 0, 1, \ldots, m-1,$$

which proves the desired result. **Q.E.D.**

B.2 WEIGHTED SUP-NORM CONTRACTIONS

In this section, we will focus on contraction mappings within a specialized context that is particularly important in DP. Let X be a set (typically the state space in DP), and let $v : X \mapsto \Re$ be a positive-valued function,

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

Let $B(X)$ denote the set of all functions $J : X \mapsto \Re$ such that $J(x)/v(x)$ is bounded as x ranges over X. We define a norm on $B(X)$, called the *weighted sup-norm*, by

$$\|J\| = \sup_{x\in X} \frac{|J(x)|}{v(x)}. \tag{B.3}$$

It is easily verified that $\|\cdot\|$ thus defined has the required properties for being a norm. Furthermore, $B(X)$ *is complete under this norm*. To see this, consider a Cauchy sequence $\{J_k\} \subset B(X)$, and note that $\|J_m - J_n\| \to 0$ as $m, n \to \infty$ implies that for all $x \in X$, $\{J_k(x)\}$ is a Cauchy sequence of real numbers, so it converges to some $J^*(x)$. We will show that $J^* \in B(X)$ and that $\|J_k - J^*\| \to 0$. To this end, it will be sufficient to show that given any $\epsilon > 0$, there exists a K such that

$$\frac{|J_k(x) - J^*(x)|}{v(x)} \le \epsilon, \qquad \forall\ x \in X,\ k \ge K.$$

This will imply that

$$\sup_{x\in X} \frac{|J^*(x)|}{v(x)} \le \epsilon + \|J_k\|, \qquad \forall\ k \ge K,$$

so that $J^* \in B(X)$, and will also imply that $\|J_k - J^*\| \le \epsilon$, so that $\|J_k - J^*\| \to 0$. Assume the contrary, i.e., that there exists an $\epsilon > 0$ and a subsequence $\{x_{m_1}, x_{m_2}, \ldots\} \subset X$ such that $m_i < m_{i+1}$ and

$$\epsilon < \frac{\big|J_{m_i}(x_{m_i}) - J^*(x_{m_i})\big|}{v(x_{m_i})}, \qquad \forall\ i \ge 1.$$

The right-hand side above is less or equal to

$$\frac{\big|J_{m_i}(x_{m_i}) - J_n(x_{m_i})\big|}{v(x_{m_i})} + \frac{\big|J_n(x_{m_i}) - J^*(x_{m_i})\big|}{v(x_{m_i})}, \qquad \forall\ n \ge 1,\ i \ge 1.$$

The first term in the above sum is less than $\epsilon/2$ for i and n larger than some threshold; fixing i and letting n be sufficiently large, the second term can also be made less than $\epsilon/2$, so the sum is made less than ϵ - a contradiction. In conclusion, the space $B(X)$ is complete, so the fixed point results of Props. B.1 and B.2 apply.

In our discussions, we will always assume that $B(X)$ is equipped with the weighted sup-norm above, where the weight function v will be clear from the context. There will be frequent occasions where the norm will be unweighted, i.e., $v(x) \equiv 1$ and $\|J\| = \max_{x\in X} \big|J(x)\big|$, in which case we will explicitly state so.

Finite-Dimensional Cases

Let us now focus on the finite-dimensional case $X = \{1,\ldots,n\}$. Consider a linear mapping $F : \Re^n \mapsto \Re^n$ of the form

$$Fy = b + Ay,$$

where A is an $n \times n$ matrix with components a_{ij}, and b is a vector in $\Re^n$ (cf. Example B.1). Then it can be shown (see the following proposition) that F *is a contraction with respect to the weighted sup-norm* $\|y\| = \max_{i=1,\ldots,n} |y_i|/v(i)$ *if and only if*

$$\frac{\sum_{j=1}^n |a_{ij}|\, v(j)}{v(i)} < 1, \qquad i = 1,\ldots,n.$$

Let us also denote by $|A|$ the matrix whose components are the absolute values of the components of A and let $\sigma\big(|A|\big)$ denote the spectral radius of $|A|$. Then it can be shown that F *is a contraction with respect to some weighted sup-norm if and only if* $\sigma\big(|A|\big) < 1$. A proof of this may be found in [BeT89], Ch. 2, Cor. 6.2. Thus any substochastic matrix P ($p_{ij} \geq 0$ for all i,j, and $\sum_{j=1}^n p_{ij} \leq 1$, for all i) is a contraction with respect to some weighted sup-norm if and only if $\sigma(P) < 1$.

Finally, let us consider a nonlinear mapping $F : \Re^n \mapsto \Re^n$ that has the property

$$|Fy - Fz| \leq P\,|y - z|, \qquad \forall\ y,z \in \Re^n,$$

for some matrix P with nonnegative components and $\sigma(P) < 1$. Here, we generically denote by $|w|$ the vector whose components are the absolute values of the components of w, and the inequality is componentwise. Then we claim that F is a contraction with respect to some weighted sup-norm. To see this note that by the preceding discussion, P is a contraction with respect to some weighted sup-norm $\|y\| = \max_{i=1,\ldots,n} |y_i|/v(i)$, and we have

$$\frac{\big(|Fy - Fz|\big)(i)}{v(i)} \leq \frac{\big(P\,|y - z|\big)(i)}{v(i)} \leq \alpha\,\|y - z\|, \qquad \forall\ i = 1,\ldots,n,$$

for some $\alpha \in (0,1)$, where $\big(|Fy - Fz|\big)(i)$ and $\big(P\,|y-z|\big)(i)$ are the ith components of the vectors $|Fy - Fz|$ and $P\,|y-z|$, respectively. Thus, F is a contraction with respect to $\|\cdot\|$. For additional discussion of linear and nonlinear contraction mapping properties and characterizations such as the one above, see the book [OrR70].

Linear Mappings on Countable Spaces

The case where X is countable (or, as a special case, finite) is frequently encountered in DP. The following proposition provides some useful criteria for verifying the contraction property of mappings that are either linear or are obtained via a parametric minimization of other contraction mappings.

Proposition B.3: Let $X = \{1, 2, \ldots\}$.

(a) Let $F : B(X) \mapsto B(X)$ be a linear mapping of the form

$$(FJ)(i) = b_i + \sum_{j\in X} a_{ij} J(j), \qquad i \in X,$$

where b_i and a_{ij} are some scalars. Then F is a contraction with modulus ρ with respect to the weighted sup-norm (B.3) if and only if

$$\frac{\sum_{j\in X} |a_{ij}|\, v(j)}{v(i)} \le \rho, \qquad i \in X. \tag{B.4}$$

(b) Let $F : B(X) \mapsto B(X)$ be a mapping of the form

$$(FJ)(i) = \inf_{\mu\in M} (F_\mu J)(i), \qquad i \in X,$$

where M is parameter set, and for each $\mu \in M$, F_μ is a contraction mapping from $B(X)$ to $B(X)$ with modulus ρ. Then F is a contraction mapping with modulus ρ.

Proof: (a) Assume that Eq. (B.4) holds. For any $J, J' \in B(X)$, we have

$$\begin{aligned}
\|FJ - FJ'\| &= \sup_{i\in X} \frac{\Big|\sum_{j\in X} a_{ij}\big(J(j) - J'(j)\big)\Big|}{v(i)} \\
&\le \sup_{i\in X} \frac{\sum_{j\in X} \big|a_{ij}\big|\, v(j) \Big(\big|J(j) - J'(j)\big|/v(j)\Big)}{v(i)} \\
&\le \sup_{i\in X} \frac{\sum_{j\in X} \big|a_{ij}\big|\, v(j)}{v(i)} \,\big\|J - J'\big\|
\end{aligned}$$

$$\leq \rho \, \|J - J'\|,$$

where the last inequality follows from the hypothesis.

Conversely, arguing by contradiction, let's assume that Eq. (B.4) is violated for some $i \in X$. Define $J(j) = v(j) \operatorname{sgn}(a_{ij})$ and $J'(j) = 0$ for all $j \in X$. Then we have $\|J - J'\| = \|J\| = 1$, and

$$\frac{\big|(FJ)(i) - (FJ')(i)\big|}{v(i)} = \frac{\sum_{j \in X} |a_{ij}| \, v(j)}{v(i)} > \rho = \rho \, \|J - J'\|,$$

showing that F is not a contraction of modulus ρ.

(b) Since F_μ is a contraction of modulus ρ, we have for any $J, J' \in B(X)$,

$$\frac{(F_\mu J)(i)}{v(i)} \leq \frac{(F_\mu J')(i)}{v(i)} + \rho \, \|J - J'\|, \qquad i \in X,$$

so by taking the infimum over $\mu \in M$,

$$\frac{(FJ)(i)}{v(i)} \leq \frac{(FJ')(i)}{v(i)} + \rho \, \|J - J'\|, \qquad i \in X.$$

Reversing the roles of J and J', we obtain

$$\frac{\big|(FJ)(i) - (FJ')(i)\big|}{v(i)} \leq \rho \, \|J - J'\|, \qquad i \in X,$$

and by taking the supremum over i, the contraction property of F is proved. **Q.E.D.**

The preceding proposition assumes that $FJ \in B(X)$ for all $J \in B(X)$. The following proposition provides conditions, particularly relevant to the DP context, which imply this assumption.

Proposition B.4: Let $X = \{1, 2, \ldots\}$, let M be a parameter set, and for each $\mu \in M$, let F_μ be a linear mapping of the form

$$(F_\mu J)(i) = b_i(\mu) + \sum_{j \in X} a_{ij}(\mu) \, J(j), \qquad i \in X.$$

(a) We have $F_\mu J \in B(X)$ for all $J \in B(X)$ provided $b(\mu) \in B(X)$ and $V(\mu) \in B(X)$, where

$$b(\mu) = \big\{b_1(\mu), b_2(\mu), \ldots\big\}, \qquad V(\mu) = \big\{V_1(\mu), V_2(\mu), \ldots\big\},$$

with

$$V_i(\mu) = \sum_{j \in X} \big|a_{ij}(\mu)\big| \, v(j), \qquad i \in X.$$

(b) Consider the mapping F

$$(FJ)(i) = \inf_{\mu \in M} (F_\mu J)(i), \qquad i \in X.$$

We have $FJ \in B(X)$ for all $J \in B(X)$, provided $b \in B(X)$ and $V \in B(X)$, where

$$b = \big\{b_1, b_2, \ldots\big\}, \qquad V = \{V_1, V_2, \ldots\},$$

with $b_i = \sup_{\mu \in M} b_i(\mu)$ and $V_i = \sup_{\mu \in M} V_i(\mu)$.

Proof: (a) For all $\mu \in M$, $J \in B(X)$ and $i \in X$, we have

$$\begin{aligned}(F_\mu J)(i) &\leq \big|b_i(\mu)\big| + \sum_{j \in X} \big|a_{ij}(\mu)\big| \, \big|J(j)/v(j)\big| \, v(j) \\ &\leq \big|b_i(\mu)\big| + \|J\| \sum_{j \in X} \big|a_{ij}(\mu)\big| \, v(j) \\ &= \big|b_i(\mu)\big| + \|J\| \, V_i(\mu),\end{aligned}$$

and similarly $(F_\mu J)(i) \geq -\big|b_i(\mu)\big| - \|J\| \, V_i(\mu)$. Thus

$$\big|(F_\mu J)(i)\big| \leq \big|b_i(\mu)\big| + \|J\| \, V_i(\mu), \qquad i \in X.$$

By dividing this inequality with $v(i)$ and by taking the supremum over $i \in X$, we obtain

$$\|F_\mu J\| \leq \|b_\mu\| + \|J\| \, \|V_\mu\| < \infty.$$

(b) By doing the same as in (a), but after first taking the infimum of $(F_\mu J)(i)$ over μ, we obtain

$$\|FJ\| \leq \|b\| + \|J\| \, \|V\| < \infty.$$

Q.E.D.

APPENDIX C:

Measure Theoretic Issues

A general theory of stochastic dynamic programming must deal with the formidable mathematical questions that arise from the presence of uncountable probability spaces. The purpose of this appendix is to motivate the theory and to provide some mathematical background to the extent needed for the development of Chapter 5. The research monograph by Bertsekas and Shreve [BeS78] (freely available from the internet), contains a detailed development of mathematical background and terminology on Borel spaces and related subjects. We will explore here the main questions by means of a simple two-stage example described in Section C.1. In Section C.2, we develop a framework, based on universally measurable policies, for the rigorous mathematical development of the standard DP results for this example and for more general finite horizon models.

C.1 A TWO-STAGE EXAMPLE

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

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

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

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

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

The optimal cost is

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

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

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

A policy π is *optimal* if

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

The DP Algorithm

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

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

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

and assuming that

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

the results we expect to be able to prove are:

R.1: There holds

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

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

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

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

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

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

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

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

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

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

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

(c) g must be such that the function

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

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

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

Countable Space Problems

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

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

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

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

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

Then

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

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

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

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

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

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

We have thus shown that when the measure $p(dx_1 \mid x_0, u_0)$ has countable support, g is bounded either above or below, and $J_0(x_0) > -\infty$ for all x_0 and $J_1(x_1) > -\infty$ for all x_1, the derivation of Eq. (C.4) is valid and proves that the DP algorithm produces the optimal cost function J^* (cf.

property R.1). † A similar argument proves the existence of an ϵ-optimal policy (cf. R.2); it uses the ϵ-optimal selection (C.5) for the second stage and a similar ϵ-optimal selection for the first stage, i.e., the existence of a $\bar{\mu}_0 : \Re \mapsto \Re$ such that

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

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

Approaches for Uncountable Space Problems

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

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

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

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

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

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

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

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

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

C.2 RESOLUTION OF THE MEASURABILITY ISSUES

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

Borel Spaces and Analytic Sets

Given a topological space Y, we denote by $\mathcal{B}_Y$ the σ-algebra generated by the open subsets of Y, and refer to the members of $\mathcal{B}_Y$ as the *Borel subsets* of Y. A topological space Y is a *Borel space* if it is homeomorphic to a Borel subset of a complete separable metric space. The concept of Borel space is quite broad, containing any "reasonable" subset of n-dimensional Euclidean space. Any Borel subset of a Borel space is again a Borel space, as is any homeomorphic image of a Borel space and any finite or countable

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

Cartesian product of Borel spaces. Let Y and Z be Borel spaces, and consider a function $h : Y \mapsto Z$. We say that h is *Borel measurable* if $h^{-1}(B) \in \mathcal{B}_Y$ for every $B \in \mathcal{B}_Z$.

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

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

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

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

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

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

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

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

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

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

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

Lower Semianalytic Functions

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

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

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

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

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

is lower semianalytic.

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

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

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

Universal Measurability

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

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

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

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

Integration of Lower Semianalytic Functions

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

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

† We use here a definition of integral of an extended real-valued function that is always defined as an extended real number (see also Appendix A). In particular. for a probability measure p, the integral of an extended real-valued function f, with positive and negative parts f^+ and f^-, is defined as

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

where we adopts the rule $\infty - \infty = \infty$ for the case where $\int f^+ dp = \infty$ and $\int f^- dp = \infty$. With this expanded definition, the integral of an extended real-valued function is always defined as an extended real number (consistently also with Appendix A).

Proposition C.4: Let Y and Z be Borel spaces, and let $q(dz \mid y)$ be a stochastic kernel on Z given Y. Let also $h : Y \times Z \mapsto [-\infty, \infty]$ be a function.

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

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

is lower semianalytic.

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

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

is universally measurable.

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

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

is a lower semianalytic function of (x_0, u_0) (cf. Prop. C.4), and in view of Prop. C.1, the same is true of the function J_0 given by the DP Eq. (C.3), which is the partial minimum over u_0 of the expression (C.7). Thus, with lower semianalytic g and Borel measurable p, the integrals appearing in the DP algorithm make sense.

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

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

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

Universally Measurable Selection

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

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

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

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

and let

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

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

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

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

Universal Measurability Framework: A Summary

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

(a) *The stage cost function g is lower semianalytic*; this is needed to show that the function J_1 of the DP Eq. (C.2) is lower semianalytic and hence also universally measurable (cf. Prop. C.1). The more "natural" Borel measurability assumption on g implies lower analyticity of g, but will not keep the functions J_1 and J_0 produced by the DP algorithm within the domain of Borel measurability. This is because the partial minimum operation on Borel measurable functions takes us outside that domain (cf. Prop. C.1).

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

(c) *The control functions μ_0 and μ_1 are allowed to be universally measurable, and we have $J_0(x_0) > -\infty$ for all x_0 and $J_1(x_1) > -\infty$ for all x_1*. This is needed in order for the calculation of Eq. (C.4) to go through (using the measurable selection property of Prop. C.5), and show that the DP algorithm produces the optimal cost function (cf. R.1). It is also needed (using again Prop. C.5) in order to show the associated existence of solutions results (cf. R.2 and R.3).

Extension to General Finite-Horizon DP

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

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

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

DP-like backwards recursion (see [BeS78], Section 8.1):

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

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

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

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

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

The DP algorithm is given by

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

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

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

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

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

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

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

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

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

Then, we can show by induction that

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

and in particular, for $k = 0$,

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

and hence also

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

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

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

APPENDIX D:

Solutions of Exercises

CHAPTER 1

1.1 (Multistep Contraction Mappings)

By the contraction property of $T_{\mu_0}, \ldots, T_{\mu_{m-1}}$, we have for all $J, J' \in B(X)$,

$$\begin{aligned}
\|\overline{T}_\nu J - \overline{T}_\nu J'\| &= \|T_{\mu_0} \cdots T_{\mu_{m-1}} J - T_{\mu_0} \cdots T_{\mu_{m-1}} J'\| \\
&\leq \alpha \|T_{\mu_1} \cdots T_{\mu_{m-1}} J - T_{\mu_1} \cdots T_{\mu_{m-1}} J'\| \\
&\leq \alpha^2 \|T_{\mu_2} \cdots T_{\mu_{m-1}} J - T_{\mu_2} \cdots T_{\mu_{m-1}} J'\| \\
&\;\;\vdots \\
&\leq \alpha^m \|J - J'\|,
\end{aligned}$$

thus showing Eq. (1.26).

We have from Eq. (1.26)

$$(T_{\mu_0} \cdots T_{\mu_{m-1}} J)(x) \leq (T_{\mu_0} \cdots T_{\mu_{m-1}} J')(x) + \alpha^m \|J - J'\|\, v(x), \qquad \forall\, x \in X,$$

and by taking infimum of both sides over $(T_{\mu_0} \cdots T_{\mu_{m-1}}) \in \mathcal{M}_m$ and dividing by $v(x)$, we obtain

$$\frac{(\overline{T}J)(x) - (\overline{T}J')(x)}{v(x)} \leq \alpha^m \|J - J'\|, \qquad \forall\, x \in X.$$

Similarly

$$\frac{(\overline{T}J')(x) - (\overline{T}J)(x)}{v(x)} \leq \alpha^m \|J - J'\|, \qquad \forall\, x \in X,$$

and by combining the last two relations and taking supremum over $x \in X$, Eq. (1.27) follows.

1.2 (State-Dependent Weighted Multistep Mappings [YuB12])

By the contraction property of T_μ, we have for all $J, J' \in B(X)$ and $x \in X$,

$$\begin{aligned}
\frac{\left|(T_\mu^{(w)}J)(x) - (T_\mu^{(w)}J')(x)\right|}{v(x)} &= \frac{\left|\sum_{\ell=1}^\infty w_\ell(x)(T_\mu^\ell J)(x) - \sum_{\ell=1}^\infty w_\ell(x)(T_\mu^\ell J')(x)\right|}{v(x)} \\
&\le \sum_{\ell=1}^\infty w_\ell(x)\|T_\mu^\ell J - T_\mu^\ell J'\| \\
&\le \left(\sum_{\ell=1}^\infty w_\ell(x)\alpha^\ell\right)\|J - J'\|,
\end{aligned}$$

showing the contraction property of $T_\mu^{(w)}$.

Let J_μ be the fixed point of T_μ. We have for all $x \in X$, by using the relation $(T_\mu^\ell J_\mu)(x) = J_\mu(x)$,

$$(T_\mu^{(w)}J_\mu)(x) = \sum_{\ell=1}^\infty w_\ell(x)\big(T_\mu^\ell J_\mu\big)(x) = \left(\sum_{\ell=1}^\infty w_\ell(x)\right)J_\mu(x) = J_\mu(x),$$

so J_μ is the fixed point of $T_\mu^{(w)}$ [which is unique since $T_\mu^{(w)}$ is a contraction].

CHAPTER 2

2.1 (Periodic Policies)

(a) Let us define

$$J_0 = \lim_{k\to\infty} \overline{T}_\nu^k \bar{J}, \quad J_1 = \lim_{k\to\infty} \overline{T}_\nu^k (T_{\mu_0}\bar{J}), \quad \dots \quad J_{m-2} = \lim_{k\to\infty} \overline{T}_\nu^k (T_{\mu_0}\cdots T_{\mu_{m-2}}\bar{J}).$$

Since $\overline{T}_\nu$ is a contraction mapping, $J_0, \dots, J_{m-1}$ are all equal to the unique fixed point of $\overline{T}_\nu$. Since $J_0, \dots, J_{m-1}$ are all equal, they are also equal to J_π (by the definition of J_π). Thus J_π is the unique fixed point of $\overline{T}_\nu$.

(b) Follow the hint.

2.2 (Totally Asynchronous Convergence Theorem for Time-Varying Maps)

A straightforward replication of the proof of Prop. 2.6.1.

2.3 (Nonmonotonic-Contractive Models – Fixed Points of Concave Sup-Norm Contractions)

The analysis of Sections 2.6.1 and 2.6.3 does not require monotonicity of the mapping T_μ given by

$$(T_\mu J)(x) = F\big(x, \mu(x)\big) - J'\mu(x).$$

2.4 (Discounted Problems with Unbounded Cost per Stage)

We have

$$\frac{\big|(T_\mu J)(x)\big|}{v(x)} \le \frac{G_x}{v(x)} + \alpha \sum_{y\in X} \frac{p_{xy}\big(\mu(x)\big)\, v(y)}{v(x)} \frac{|J(y)|}{v(y)}, \qquad \forall\ x \in X,\ \mu \in \mathcal{M},$$

from which, using assumptions (1) and (2),

$$\frac{\big|(T_\mu J)(x)\big|}{v(x)} \le \|G\| + \alpha\|V\|\,\|J\|, \qquad \forall\ x \in X,\ \mu \in \mathcal{M}.$$

A similar argument shows that

$$\frac{\big|(TJ)(x)\big|}{v(x)} \le \|G\| + \alpha\|V\|\,\|J\|, \qquad \forall\ x \in X.$$

It follows that $T_\mu J \in B(X)$ and $TJ \in B(X)$ if $J \in B(X)$.

For any $J, J' \in B(X)$ and $\mu \in \mathcal{M}$, we have

$$\begin{aligned}
\|T_\mu J - T_\mu J'\| &= \sup_{x\in X} \frac{\Big|\alpha \sum_{y\in X} p_{xy}\big(\mu(x)\big)\big(J(y) - J'(y)\big)\Big|}{v(x)} \\
&\le \sup_{x\in X} \frac{\Big|\alpha \sum_{y\in X} p_{xy}\big(\mu(x)\big)v(y)\big(|J(y) - J'(y)|/v(y)\big)\Big|}{v(x)} \\
&\le \sup_{x\in X} \alpha \frac{\Big|\sum_{y\in X} p_{xy}\big(\mu(x)\big)v(y)\Big|}{v(x)} \|J - J'\| \\
&\le \alpha\|J - J'\|,
\end{aligned}$$

where the last inequality follows from assumption (3). Hence T_μ is a contraction of modulus α.

To show that T is a contraction, we note that

$$\frac{(T_\mu J)(x)}{v(x)} \le \frac{(T_\mu J')(x)}{v(x)} + \alpha\|J - J'\|, \qquad x \in X,\ \mu \in \mathcal{M},$$

so by taking infimum over $\mu \in \mathcal{M}$, we obtain

$$\frac{(TJ)(x)}{v(x)} \le \frac{(TJ')(x)}{v(x)} + \alpha\|J - J'\|, \qquad x \in X.$$

Similarly,

$$\frac{(TJ')(x)}{v(x)} \le \frac{(TJ)(x)}{v(x)} + \alpha\|J - J'\|, \qquad x \in X,$$

and by combining the last two relations the contraction property of T follows.

2.5 (Solution by Math. Programming)

If $J \leq TJ$, by monotonicity we have $J \leq \lim_{k\to\infty} T^k J = J^*$. Any feasible solution z of the optimization problem satisfies $z_i \leq H(i,u,z)$ for all $i = 1,\dots,n$ and $u \in U(i)$, so that $z \leq Tz$. It follows that $z \leq J^*$, which implies that J^* is an optimal solution of the optimization problem.

2.6 (Convergence of Nonexpansive Monotone Fixed Point Iterations)

For any $c > 0$, let $V_k = T^k(J^* + c\,v)$ for $k \geq 1$, and note that $J^* = T^k J^* \leq V_k$. From Eq. (2.80), we have

$$H(x,u,J^* + c\,v) \leq H(x,u,J^*) + c\,v(x), \qquad x \in X,\ u \in U(x),$$

and by taking the minimum over $u \in U(x)$, we obtain $T(J^* + c\,v) \leq J^* + c\,v$, i.e., $V_1 \leq V_0$. From this and the monotonicity of T it follows that $\big\{V_k(x)\big\}$ is monotonically nonincreasing, and converges to some scalar $\bar V(x) \geq J^*(x)$ for each $x \in X$. Moreover, the corresponding function $\bar V$ is in $B(X)$, since $V_0 \geq \bar V \geq J^*$, and also satisfies $\|V_k - \bar V\| \to 0$ (since X is finite). From Eq. (2.80), we have $\|TV_k - T\bar V\| \leq \|V_k - \bar V\|$, so $\|TV_k - T\bar V\| \to 0$ which together with the fact $TV_k = V_{k+1} \to \bar V$, implies that $\bar V = T\bar V$. Thus $\bar V = J^*$ by the uniqueness of the fixed point of T, and it follows that $\{V_k\}$ converges monotonically to J^* from above.

Similarly, define $W_k = T^k(J^* - c\,v)$, and by an argument symmetric to the above, $\{W_k\}$ converges monotonically to J^* from below. Now let $c = \|J - J^*\|$ in the definition of V_k and W_k. Then $J^* - c\,v \leq J_0 = J \leq J^* + c\,v$, so by the monotonicity of T, $W_k \leq T^k J \leq V_k$ as well as $W_k \leq J^* \leq V_k$ for all k. Therefore

$$\frac{\big|(T^k J)(x) - J^*(x)\big|}{v(x)} \leq \frac{\big|W_k(x) - V_k(x)\big|}{v(x)} \leq \|W_k - V_k\|, \qquad \forall\ x \in X.$$

Since $\|W_k - V_k\| \leq \|W_k - J^*\| + \|V_k - J^*\| \to 0$, the conclusion follows.

CHAPTER 3

3.1 (Blackmailer's Dilemma)

(a) Clearly T_μ is a sup-norm contraction with modulus $1 - \mu(1)^2$. Hence J_μ is the unique fixed point of T_μ and we have

$$J_\mu(1) = (T_\mu J_\mu)(1) = -\mu(1) + \big(1 - \mu(1)^2\big) J_\mu(1),$$

which yields $J_\mu(1) = -1/\mu(1)$. The mapping T is given by

$$(TJ)(1) = \inf_{0<u\leq 1} \big\{ -u + (1-u^2)J(1)\big\},$$

and $J \in \Re$ is a fixed point of T if and only if

$$0 = \inf_{0<u\leq 1} \Big\{ - \big(u + u^2 J(1)\big)\Big\}.$$

However, it can be seen that this equation has no solution. Here parts (b) and (d) of Assumption 3.2.1 are violated.

(b) Here T_μ is again a sup-norm contraction with modulus $1-\mu(1)^2$. For J_μ, the unique fixed point of T_μ, we have

$$J_\mu(1) = (T_\mu J_\mu)(1) = -\big(1-\mu(1)\big)\mu(1) + \big(1-\mu(1)\big) J_\mu(1),$$

which yields $J_\mu(1) = -1 + \mu(1)$. Hence $J^* = -1$, but there is no optimal μ. The mapping T is given by

$$(TJ)(1) = \inf_{0<u\leq 1} \Big\{ - u + u^2 + (1-u)J(1)\Big\},$$

and $J \in \Re$ is a fixed point of T if and only if

$$0 = \inf_{0<u\leq 1} \Big\{ - u + u^2 - uJ(1)\Big\}.$$

It can be verified that the set of fixed points of T within $\Re$ is $\{J \mid J \leq -1\}$. Here part (d) of Assumption 3.2.1 is violated.

(c) For the policy $\overline{\mu}$ that chooses $\overline{\mu}(1) = 0$, we have

$$(T_{\overline{\mu}} J)(1) = c + J(1),$$

and $\overline{\mu}$ is $\Re$-irregular since $\lim_{k\to\infty} T^k_{\overline{\mu}} J$ either does not belong to $\Re$ or depends on J. Moreover, the mapping T is given by

$$(TJ)(1) = \min \left\{ c + J(1),\ \inf_{0<u\leq 1} \Big\{ - u + u^2 + (1-u)J(1)\Big\} \right\}.$$

When $c > 0$, we have $J_{\overline{\mu}}(1) = \lim_{k\to\infty} (T^k_{\overline{\mu}} \bar J)(1) = \infty$. It can be verified that there is no optimal policy, and the set of fixed points of T within $\Re$ is $\{J \mid J \leq -1\}$. Here part (d) of Assumption 3.2.1 is violated.

When $c = 0$, we have $J_{\overline{\mu}}(1) = \lim_{k\to\infty} (T^k_{\overline{\mu}} \bar J)(1) = 0$. Again it can be verified that there is no optimal policy, and the set of fixed points of T within $\Re$ is $\{J \mid J \leq -1\}$. Here part (c) of Assumption 3.2.1 is violated.

When $c < 0$, we have $J_{\overline{\mu}}(1) = \lim_{k\to\infty} (T^k_{\overline{\mu}} \bar J)(1) = -\infty$, and the $\Re$-irregular policy $\overline{\mu}$ is optimal. The mapping T has no fixed point within $\Re$. Here parts (c) and (d) of Assumption 3.2.1 are violated.

3.2 (Equivalent Semicontractive Conditions)

Let the assumptions of Prop. 3.1.1 hold, and let μ^* be the S-regular policy that is optimal. Then condition (1) implies that $J^* = J_{\mu^*} \in S$ and $J^* = T_{\mu^*} J^* \geq TJ^*$, while condition (2) implies that there exists an S-regular policy μ such that $T_\mu J^* = TJ^*$.

Conversely, assume that $J^* \in S$, $TJ^* \leq J^*$, and there exists an S-regular policy μ such that $T_\mu J^* = TJ^*$. Then we have $T_\mu J^* = TJ^* \leq J^*$. Hence $T^k_\mu J^* \leq J^*$ for all k, and by taking the limit as $k \to \infty$, we obtain $J_\mu \leq J^*$. Hence the S-regular policy μ is optimal, and both conditions of Prop. 3.1.1 hold.

3.3

The mapping H here is

$$H(x,u,J) = \begin{cases} b & \text{if } x=1,\ u=0, \\ a+J(2) & \text{if } x=1,\ u=2, \\ a+J(1) & \text{if } x=2,\ u=1. \end{cases}$$

The Bellman equation is given by

$$J(1) = \min\big\{b, a+J(2)\big\}, \qquad J(2) = a+J(1).$$

There are two policies:

$$\mu: \text{ where } \mu(1)=0, \text{ corresponding to the path } 2\to1\to0,$$

$$\overline{\mu}: \text{ where } \overline{\mu}(1)=2, \text{ corresponding to the cycle } 1\to2\to1.$$

The case where $S=\Re^2$ has been discussed in Section 3.1.2. Here μ is S-regular, as can be seen from the form of T_μ,

$$(T_\mu J)(1)=b, \qquad (T_\mu J)(2)=a+J(1),$$

but $\overline{\mu}$ is S-irregular, as can be seen from the form of $T_{\overline{\mu}}$,

$$(T_{\overline{\mu}}J)(1)=a+J(2), \qquad (T_{\overline{\mu}}J)(2)=a+J(1).$$

Briefly there are four cases of interest:

(1) $\alpha>0$: Here Prop. 3.1.1 applies.

(2) $\alpha=0$ and $b\le0$: Here Prop. 3.1.1 applies.

(3) $\alpha=0$ and $b>0$: Here Prop. 3.1.1 does not apply because the S-regular policy μ is not optimal.

(4) $\alpha<0$: Here Prop. 3.1.1 does not apply because the S-regular policy μ is not optimal.

Consider now the case where $S=[-\infty,\infty)\times[-\infty,\infty)$. Then μ is S-regular in all cases (1)-(4), but $\overline{\mu}$ is S-irregular only in cases (1)-(3), and it is S-regular in case (4) because $J_{\overline{\mu}}(1)=J_{\overline{\mu}}(2)=-\infty$ and

$$\lim_{k\to\infty}(T_{\overline{\mu}}^kJ)(1) = \lim_{k\to\infty}(T_{\overline{\mu}}^kJ)(1) = -\infty, \qquad \forall\ J\in S,$$

while $J_{\overline{\mu}}$ is the unique fixed point of T within S. In cases (1) and (2), Prop. 3.1.1 applies, because the S-regular policy μ is optimal. In case (3), Prop. 3.1.1 does not apply because the S-regular policy μ is not optimal. Finally, in case (4), contrary to the case $S=\Re^2$, Prop. 3.1.1 applies, because the policy $\overline{\mu}$ is optimal and also S-regular. Case (3) cannot be analyzed with the aid of Props. 3.1.1, 3.1.2, or 3.2.1.

3.4 (Changing $\bar{J}$)

(a) By the definition of S-regular policy, we have $T^k J \to J_\mu$ for all S-regular μ and $J \in S$. Thus, changing $\bar{J}$ to $J \in S$ leaves the cost function of all S-regular policies unchanged.

(b) Here

$$H(x,u,J) = \begin{cases} b & \text{if } x = 1,\ u = 0, \\ J(2) & \text{if } x = 1,\ u = 2, \\ J(1) & \text{if } x = 2,\ u = 1. \end{cases}$$

When $\bar{J} = 0$, the $\Re^2$-regular policy is optimal and $J^* = b\,e$, as shown in Section 3.1.2. When $\bar{J} = r\,e$, the cost function of the $\Re^2$-regular policy μ [$\mu(1) = 0$] continues to be

$$J_\mu(1) = J_\mu(2) = b,$$

while the cost function of the $\Re^2$-irregular policy $\overline{\mu}$ [$\overline{\mu}(1) = 2$] is

$$J_{\overline{\mu}}(1) = J_{\overline{\mu}}(2) = r.$$

For $r \le b$, the $\Re^2$-irregular policy is optimal, but $J^* = b\,e$ continues to be the optimal cost over just the $\Re^2$-regular policies (there is only one in this example).

3.5 (Alternative Semicontractive Conditions)

We will show that conditions (1) and (2) imply that $J^* = TJ^*$, and the result will follow from Prop. 3.1.2. Assume to obtain a contradiction, that $J^* \neq TJ^*$. Then $J^* \ge TJ^*$, as can be seen from the relations

$$J^* = J_{\mu^*} = T_{\mu^*} J_{\mu^*} \ge T J_{\mu^*} = TJ^*,$$

where μ^* is an optimal S-regular policy. Thus the relation $J^* \neq TJ^*$ implies that there exists $\overline{\mu}$ and $x \in X$ such that

$$J^*(x) \ge (T_{\overline{\mu}} J^*)(x), \qquad \forall\, x \in X,$$

with strict inequality for some x [note here that we can choose $\overline{\mu}(x) = \mu^*(x)$ for all x such that $J^*(x) = (TJ^*)(x)$, and we can choose $\overline{\mu}(x)$ to satisfy $J^*(x) > (T_{\overline{\mu}} J^*)(x)$ for all other x]. If $\overline{\mu}$ were S-regular, we would have

$$J^* \ge T_{\overline{\mu}} J^* \ge \lim_{k\to\infty} T_{\overline{\mu}}^k J^* = J_{\overline{\mu}},$$

with strict inequality for some $x \in X$, which is impossible. Hence $\overline{\mu}$ is S-irregular, which contradicts condition (2).

3.6 (Convergence of PI)

We have

$$J_{\mu^k} \geq TJ_{\mu^k} \geq J_{\mu^{k+1}}, \qquad k = 0, 1, \ldots. \tag{3.1}$$

Denote

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

Since for all k, we have $J_{\mu^k} \geq \hat{J} \in S$, where $\hat{J}$ is the optimal cost function over S-regular policies [cf. Assumption 3.2.1(b)]. It follows that $J_\infty \geq \hat{J}$, and by Assumption 3.2.1(a), we obtain $J_\infty \in S$. By taking the limit in Eq. (3.1), we have

$$J_\infty = \lim_{k\to\infty} TJ_{\mu^k} \geq TJ_\infty, \tag{3.2}$$

where the inequality follows from the fact $J_{\mu^k} \downarrow J_\infty$. Using also the given assumption, we have for all $x \in X$ and $u \in U(x)$,

$$H(x, u, J_\infty) = \lim_{k\to\infty} H(x, u, J_{\mu^k}) \geq \lim_{k\to\infty} (TJ_{\mu^k})(x) = J_\infty(x).$$

By taking the infimum of the left-hand side over $u \in U(x)$, we obtain $TJ_\infty \geq J_\infty$, which combined with Eq. (3.2), yields $J_\infty = TJ_\infty$. Since J^* is the unique fixed point of T within S, we obtain $J_\infty = J^*$.

CHAPTER 4

4.1 (Example of Nonexistence of an Optimal Policy Under D)

Since a cost is incurred only upon stopping, and the stopping cost is greater than -1, we have $J_\mu(x) > -1$ for all x and μ. On the other hand, starting from any state x and stopping at $x+n$ yields a cost $-1 + \frac{1}{x+n}$, so by taking n sufficiently large, we can attain a cost arbitrarily close to -1. Thus $J^*(x) = -1$ for all x, but no policy can attain this optimal cost.

4.2 (Counterexample for Optimality Condition Under D)

We have $J^*(x) = -1$ and $J_\mu(x) = 0$ for all $x \in X$. Thus μ is nonoptimal, yet attains the minimum in Bellman's equation

$$J^*(x) = \min\left\{J^*(x+1),\ -1 + \frac{1}{x}\right\}$$

for all x.

4.3 (Counterexample for Optimality Condition Under I)

The verification of $T_\mu J_\mu = TJ_\mu$ is straightforward. To show that $J^*(x) = |x|$, we first note that $|x|$ is a fixed point of T, so by Prop. 4.3.2, $J^*(x) \le |x|$. Also $(T\bar{J})(x) = |x|$ for all x, while under Assumption I, we have $J^* \ge T\bar{J}$, so $J^*(x) \ge |x|$. Hence $J^*(x) = |x|$.

4.4 (Solution by Math. Programming)

(a) Any feasible solution z of the given optimization problem satisfies $z \ge \bar{J}$ as well as $z_i \ge \inf_{u\in U(i)} H(i,u,z)$ for all $i = 1,\ldots,n$, so that $z \ge Tz$. It follows from Prop. 4.3.3 that $z \ge J^*$, which implies that J^* is an optimal solution of the given optimization problem. Also J^* is the unique optimal solution since if z is feasible and $z \ne J^*$, the inequality $z \ge J^*$ implies that $\sum_i z_i > \sum_i J^*(i)$, so z cannot be optimal.

(b) Any feasible solution z of the given optimization problem satisfies $z \le \bar{J}$ as well as $z_i \le H(i,u,z)$ for all $i = 1,\ldots,n$ and $u \in U(i)$, so that $z \le Tz$. It follows from Prop. 4.3.6 that $z \le J^*$, which implies that J^* is an optimal solution of the given optimization problem. Similar to part (a), J^* is the unique optimal solution.

4.5 (Semicontractive Discounted Problems with Unbounded Cost per Stage)

(a) See Exercise 2.4.

(b) Since all policies in $\overline{\mathcal{M}}$ are S-regular and there exists an optimal policy within $\overline{\mathcal{M}}$, it follows that Prop. 4.4.1 applies, so that J^* is the unique fixed point of T within S. Similarly, the assumption that for each $J \in S$ there exists $\mu \in \overline{\mathcal{M}}$ such that $T_\mu J = TJ$, and the structure of H and S imply that Prop. 4.4.2 applies.

4.6 (Blackmailer's Dilemma)

(a) From Exercise 3.1, the cost function of any policy μ is

$$J_\mu(1) = -\frac{1}{\mu(1)},$$

so the policy evaluation equation given in part (a) is correct. Moreover, we have $J_\mu(1) \le -1$ since $\mu(1) \in (0,1]$. The policy improvement equation is

$$\mu^{k+1}(1) \in \arg\min_{u\in(0,1]} \big\{ -u + (1-u^2)J_{\mu^k}(1)\big\}. \tag{4.1}$$

By setting the gradient of the expression within braces to 0,

$$0 = -1 - 2uJ_{\mu^k}(1),$$

we see that its unconstrained minimum is

$$u_k = -\frac{1}{2J_{\mu^k}(1)},$$

which is less or equal to $-1/2$ since $J_\mu(1) \le -1$ for all μ. Hence u_k is equal to the constrained minimum in Eq. (4.1), and we have

$$\mu^{k+1}(1) = -\frac{1}{2J_{\mu^k}(1)}.$$

(b) Follows from Props. 4.3.14 and 4.3.15.

4.7 (Counterexample for Policy Improvement Under D - Infinite State Space)

(a) The policy μ that stops at every state has cost function

$$J_\mu(x) = -1 + \frac{1}{x}, \qquad x \in X.$$

Policy improvement starting with μ yields $\overline{\mu}$ with

$$\overline{\mu}(x) \in \arg\min \left\{ J_\mu(x),\, -1 + \frac{1}{x} \right\},$$

so $\overline{\mu}(x)$ can be either to continue or to stop at every x. Let $\overline{\mu}$ be to continue at every x. Then $J_{\overline{\mu}}(x) = 0 > J_\mu(x)$ for all x. Moreover, the next policy obtained from $\overline{\mu}$ by policy improvement is μ.

(b) Follows from Props. 4.3.14 and 4.3.15.

4.8 (Counterexample for Policy Improvement Under D - Finite State Space)

(a) Essentially the same as the one of Exercise 4.7.

(b) Straightforward.

4.9 (Infinite Time Reachability [Ber72])

(a) For any policy $\pi = \{\mu_0, \mu_1, \ldots\}$, we have

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

The mapping T_μ has the property that if J takes only the two values 0 and ∞, the same is true for $T_\mu J$. It follows that $T_{\mu_0} \cdots T_{\mu_k} \bar{J}$ takes only the two values 0 and ∞, and therefore the same is true for J_π. It can be shown by induction that $T_{\mu_0} \cdots T_{\mu_k} \bar{J}$ takes the value ∞ for all $\cap_{k=0}^{\infty} X_k$. Hence the set of states X^* where J^* takes the value 0 is a subset of $\cap_{k=0}^{\infty} X_k$ for all k, and it follows that $X^* \subset \cap_{k=0}^{\infty} X_k$.

(b) This is a consequence of the fact that Assumption I holds and $\{T^k \bar{J}\}$ is monotonically nondecreasing and satisfies $T^k \bar{J} \leq J^*$ for all k.

(c) The relation $X^* \neq \cap_{k=0}^{\infty} X_k$ is equivalent to $\lim_{k\to\infty} T^k \bar{J} \neq J^*$. The compactness condition of Prop. 4.3.13 requires that the sets

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

are compact for every $x \in X$, $\lambda \in \Re$, and for all k greater than some integer $\overline{k}$. Equivalently, the sets X_k should be compact.

References

[ABB02] Abounadi, J., Bertsekas, B. P., and Borkar, V. S., 2002. "Stochastic Approximation for Non-Expansive Maps: Q-Learning Algorithms," SIAM J. on Control and Opt., Vol. 41, pp. 1-22.

[BBB08] Basu, A., Bhattacharyya, and Borkar, V., 2008. "A Learning Algorithm for Risk-Sensitive Cost," Math. of OR, Vol. 33, pp. 880-898.

[BBD10] Busoniu, L., Babuska, R., De Schutter, B., and Ernst, D., 2010. Reinforcement Learning and Dynamic Programming Using Function Approximators, CRC Press, N. Y.

[Bau78] Baudet, G. M., 1978. "Asynchronous Iterative Methods for Multiprocessors," Journal of the ACM, Vol. 25, pp. 226-244.

[BeI96] Bertsekas, D. P., and Ioffe, S., 1996. "Temporal Differences-Based Policy Iteration and Applications in Neuro-Dynamic Programming," Lab. for Info. and Decision Systems Report LIDS-P-2349, MIT.

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

[BeT89] Bertsekas, D. P., and Tsitsiklis, J. N., 1989. Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Engl. Cliffs, N. J.; may be downloaded from http://web.mit.edu/dimitrib/www/home.html

[BeT91] Bertsekas, D. P., and Tsitsiklis, J. N., 1991. "An Analysis of Stochastic Shortest Path Problems," Math. of OR, Vol. 16, pp. 580-595.

[BeT96] Bertsekas, D. P., and Tsitsiklis, J. N., 1996. Neuro-Dynamic Programming, Athena Scientific, Belmont, MA.

[BeT08] Bertsekas, D. P., and Tsitsiklis, J. N., 2008. Introduction to Probability, 2nd Ed., Athena Scientific, Belmont, MA.

[BeY07] Bertsekas, D. P., and Yu, H., 2007. "Solution of Large Systems of Equations Using Approximate Dynamic Programming Methods," Lab. for Info. and Decision Systems Report LIDS-P-2754, MIT.

[BeY09] Bertsekas, D. P., and Yu, H., 2009. "Projected Equation Methods for Approximate Solution of Large Linear Systems," J. of Computational and Applied Mathematics, Vol. 227, pp. 27-50.

[BeY10a] Bertsekas, D. P., and Yu, H., 2010. "Q-Learning and Enhanced Policy Iteration in Discounted Dynamic Programming," Lab. for Info. and Decision Systems Report LIDS-P-2831, MIT; Math. of OR, Vol. 37, 2012, pp. 66-94.

[BeY10b] Bertsekas, D. P., and Yu, H., 2010. “Asynchronous Distributed Policy Iteration in Dynamic Programming,” Proc. of Allerton Conf. on Communication, Control and Computing, Allerton Park, Ill, pp. 1368-1374.

[Ber72] Bertsekas, D. P., 1972. “Infinite Time Reachability of State Space Regions by Using Feedback Control,” IEEE Trans. Aut. Control, Vol. AC-17, pp. 604-613.

[Ber77] Bertsekas, D. P., 1977. “Monotone Mappings with Application in Dynamic Programming,” SIAM J. on Control and Opt., Vol. 15, pp. 438-464.

[Ber82] Bertsekas, D. P., 1982. “Distributed Dynamic Programming,” IEEE Trans. Aut. Control, Vol. AC-27, pp. 610-616.

[Ber83] Bertsekas, D. P., 1983. “Asynchronous Distributed Computation of Fixed Points,” Math. Programming, Vol. 27, pp. 107-120.

[Ber87] Bertsekas, D. P., 1987. Dynamic Programming: Deterministic and Stochastic Models, Prentice-Hall, Englewood Cliffs, N. J.

[Ber05a] Bertsekas, D. P., 2005. Dynamic Programming and Optimal Control, Vol. I, 3rd Edition, Athena Scientific, Belmont, MA.

[Ber05b] Bertsekas, D. P., 2005. “Dynamic Programming and Suboptimal Control: A Survey from ADP to MPC,” Fundamental Issues in Control, Special Issue for the CDC-ECC 05, European J. of Control, Vol. 11, Nos. 4-5.

[Ber09] Bertsekas, D. P., 2009. Convex Optimization Theory, Athena Scientific, Belmont, MA.

[Ber10] Bertsekas, D. P., 2010. “Williams-Baird Counterexample for Q-Factor Asynchronous Policy Iteration,”
http://web.mit.edu/dimitrib/www/Williams-Baird Counterexample.pdf

[Ber11a] Bertsekas, D. P., 2011. “Temporal Difference Methods for General Projected Equations,” IEEE Trans. Aut. Control, Vol. 56, pp. 2128-2139.

[Ber11b] Bertsekas, D. P., 2011. “λ-Policy Iteration: A Review and a New Implementation,” Lab. for Info. and Decision Systems Report LIDS-P-2874, MIT; appears in Reinforcement Learning and Approximate Dynamic Programming for Feedback Control, by F. Lewis and D. Liu (eds.), IEEE Press, 2012.

[Ber11c] Bertsekas, D. P., 2011. “Approximate Policy Iteration: A Survey and Some New Methods,” J. of Control Theory and Applications, Vol. 9, pp. 310-335.

[Ber12a] Bertsekas, D. P., 2012. Dynamic Programming and Optimal Control, Vol. II, 4th Edition: Approximate Dynamic Programming, Athena Scientific, Belmont, MA.

[Ber12b] Bertsekas, D. P., 2012. “Weighted Sup-Norm Contractions in Dynamic Programming: A Review and Some New Applications,” Lab. for Info. and Decision Systems Report LIDS-P-2884, MIT.

[Bla65] Blackwell, D., 1965. “Positive Dynamic Programming,” Proc. Fifth Berkeley Symposium Math. Statistics and Probability, pp. 415-418.

[BoM99] Borkar, V. S., Meyn, S. P., 1999. “Risk Sensitive Optimal Control: Existence and Synthesis for Models with Unbounded Cost,” SIAM J. Control and Opt., Vol. 27, pp. 192-209.

[BoM00] Borkar, V. S., Meyn, S. P., 1900. “The O.D.E. Method for Convergence of Stochastic Approximation and Reinforcement Learning,” SIAM J. Control and Opt., Vol. 38, pp. 447-469.

[BoM02] Borkar, V. S., Meyn, S. P., 2002. "Risk-Sensitive Optimal Control for Markov Decision Processes with Monotone Cost," Math. of OR, Vol. 27, pp. 192-209.

[Bor98] Borkar, V. S., 1998. "Asynchronous Stochastic Approximation," SIAM J. Control Opt., Vol. 36, pp. 840-851.

[Bor08] Borkar, V. S., 2008. Stochastic Approximation: A Dynamical Systems Viewpoint, Cambridge Univ. Press, N. Y.

[CFH07] Chang, H. S., Fu, M. C., Hu, J., Marcus, S. I., 2007. Simulation-Based Algorithms for Markov Decision Processes, Springer, N. Y.

[CaM88] Carraway, R. L., and Morin, T. L., 1988. "Theory and Applications of Generalized Dynamic Programming: An Overview," Computers and Mathematics with Applications, Vol. 16, pp. 779-788.

[CaR12] Cavus, O., and Ruszczynski, A., 2012. "Risk-Averse Control of Undiscounted Transient Markov Models," Rutgers Univ. Report, available on Optimization On-Line at http://www.optimization-online.org/

[Cao07] Cao, X. R., 2007. Stochastic Learning and Optimization: A Sensitivity-Based Approach, Springer, N. Y.

[ChM69] Chazan D., and Miranker, W., 1969. "Chaotic Relaxation," Linear Algebra and Applications, Vol. 2, pp. 199-222.

[ChS87] Chung, K.-J., and Sobel, M. J., 1987. "Discounted MDPs: Distribution Functions and Exponential Utility Maximization," SIAM J. Control and Opt., Vol. 25, pp. 49-62.

[CoM99] Coraluppi, S. P., and Marcus, S. I., 1999. "Risk-Sensitive and Minimax Control of Discrete-Time, Finite-State Markov Decision Processes," Automatica, Vol. 35, pp. 301-309.

[DeM67] Denardo, E. V., and Mitten, L. G., 1967. "Elements of Sequential Decision Processes," J. Indust. Engrg., Vol. 18, pp. 106-112.

[DeR79] Denardo, E. V., and Rothblum, U. G., 1979. "Optimal Stopping, Exponential Utility, and Linear Programming," Math. Programming, Vol. 16, pp. 228-244.

[Den67] Denardo, E. V., 1967. "Contraction Mappings in the Theory Underlying Dynamic Programming," SIAM Review, Vol. 9, pp. 165-177.

[Der70] Derman, C., 1970. Finite State Markovian Decision Processes, Academic Press, N. Y.

[DuS65] Dubins, L., and Savage, L. M., 1965. How to Gamble If You Must, McGraw-Hill, N. Y.

[DyY79] Dynkin, E. B., and Yushkevich, A. A., 1979. Controlled Markov Processes, Springer-Verlag, Berlin and New York.

[FeM97] Fernandez-Gaucherand, E., and Marcus, S. I., 1997. "Risk-Sensitive Optimal Control of Hidden Markov Models: Structural Results," IEEE Trans. Aut. Control, Vol. AC-42, pp. 1418-1422.

[FiV96] Filar, J., and Vrieze, K., 1996. Competitive Markov Decision Processes, Springer, N. Y.

[FlM95] Fleming, W. H., and McEneaney, W. M., 1995. "Risk-Sensitive Control on an Infinite Time Horizon," SIAM J. Control and Opt., Vol. 33, pp. 1881-1915.

[Gos03] Gosavi, A., 2003. Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning, Springer, N. Y.

[HCP99] Hernandez-Lerma, O., Carrasco, O., and Perez-Hernandez. 1999. "Markov Control Processes with the Expected Total Cost Criterion: Optimality, Stability, and Transient Models," Acta Appl. Math., Vol. 59, pp. 229-269.

[Hay08] Haykin, S., 2008. Neural Networks and Learning Machines, (3rd Edition), Prentice-Hall, Englewood-Cliffs, N. J.

[HeL99] Hernandez-Lerma, O., and Lasserre, J. B., 1999. 'Further Topics on Discrete-Time Markov Control Processes, Springer, N. Y.

[HeM96] Hernandez-Hernndez, D., and Marcus, S. I., 1996. "Risk Sensitive Control of Markov Processes in Countable State Space," Systems and Control Letters, Vol. 29, pp. 147-155.

[HiW05] Hinderer, K., and Waldmann, K.-H., 2005. "Algorithms for Countable State Markov Decision Models with an Absorbing Set," SIAM J. of Control and Opt., Vol. 43, pp. 2109-2131.

[HoM72] Howard, R. S., and Matheson, J. E., 1972. "Risk-Sensitive Markov Decision Processes," Management Science, Vol. 8, pp. 356-369.

[JBE94] James, M. R., Baras, J. S., Elliott, R. J., 1994. "Risk-Sensitive Control and Dynamic Games for Partially Observed Discrete-Time Nonlinear Systems," IEEE Trans. Aut. Control, Vol. AC-39, pp. 780-792.

[JaC06] James, H. W., and Collins, E. J., 2006. "An Analysis of Transient Markov Decision Processes," J. Appl. Prob., Vol. 43, pp. 603-621.

[Jac73] Jacobson, D. H., 1973. "Optimal Stochastic Linear Systems with Exponential Performance Criteria and their Relation to Deterministic Differential Games," IEEE Transactions on Automatic Control, Vol. AC-18, pp. 124-131.

[Kle68] Kleinman, D. L., 1968. "On an Iterative Technique for Riccati Equation Computations," IEEE Trans. Aut. Control, Vol. AC-13, pp. 114-115.

[Mey07] Meyn, S., 2007. Control Techniques for Complex Networks, Cambridge Univ. Press, N. Y.

[Mit64] Mitten, L. G., 1964. "Composition Principles for Synthesis of Optimal Multistage Processes," Operations Research, Vol. 12, pp. 610-619.

[Mor82] Morin, T. L., 1982. "Monotonicity and the Principle of Optimality," J. of Math. Analysis and Applications, Vol. 88, pp. 665-674.

[OrR70] Ortega, J. M., and Rheinboldt, W. C., 1970. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, N. Y.

[PaB99] Patek, S. D., and Bertsekas, D. P., 1999. "Stochastic Shortest Path Games," SIAM J. on Control and Opt., Vol. 36, pp. 804-824.

[Pal67] Pallu de la Barriere, R., 1967. Optimal Control Theory, Saunders, Phila; republished by Dover, N. Y., 1980.

[Pat01] Patek, S. D., 2001. "On Terminating Markov Decision Processes with a Risk Averse Objective Function," Automatica, Vol. 37, pp. 1379-1386.

[Pat07] Patek, S. D., 2007. "Partially Observed Stochastic Shortest Path Problems with Approximate Solution by Neuro-Dynamic Programming," IEEE Trans. on Systems, Man, and Cybernetics Part A, Vol. 37, pp. 710-720.

[Pli78] Pliska, S. R., 1978. "On the Transient Case for Markov Decision Chains with General State Spaces," in Dynamic Programming and its Applications, by M. L. Puterman (ed.), Academic Press, N. Y.

[Pow07] Powell, W. B., 2007. Approximate Dynamic Programming: Solving the Curses of Dimensionality, J. Wiley and Sons, Hoboken, N. J; 2nd ed., 2011.

[Put94] Puterman, M. L., 1994. Markovian Decision Problems, J. Wiley, N. Y.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Ros67] Rosenfeld, J., 1967. "A Case Study on Programming for Parallel Processors," Research Report RC-1864, IBM Res. Center, Yorktown Heights, N. Y.

[Rot79] Rothblum, U. G., 1979. "Iterated Successive Approximation for Sequential Decision Processes," in Stochastic Control and Optimization, by J. W. B. van Overhagen and H. C. Tijms (eds), Vrije University, Amsterdam.

[Rot84] Rothblum, U. G., 1984. "Multiplicative Markov Decision Chains," Math. of OR, Vol. 9, pp. 6-24.

[Rus10] Ruszczynski, A., 2010. "Risk-Averse Dynamic Programming for Markov Decision Processes," Math. Programming, Ser. B, Vol. 125, pp. 235-261.

[ScL12] Scherrer, B., and Lesner, B., 2012. "On the Use of Non-Stationary Policies for Stationary Infinite-Horizon Markov Decision Processes," NIPS 2012 - Neural Information Processing Systems, South Lake Tahoe, Ne.

[Sch75] Schal, M., 1975. "Conditions for Optimality in Dynamic Programming and for the Limit of n-Stage Optimal Policies to be Optimal," Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, Vol. 32, pp. 179-196.

[Sch11] Scherrer, B., 2011. "Performance Bounds for Lambda Policy Iteration and Application to the Game of Tetris," Report RR-6348, INRIA, France; to appear in J. of Machine Learning Research.

[Sch12] Scherrer, B., 2012. "On the Use of Non-Stationary Policies for Infinite-Horizon Discounted Markov Decision Processes," INRIA Lorraine Report, France.

[Sha53] Shapley, L. S., 1953. "Stochastic Games," Proc. Nat. Acad. Sci. U.S.A., Vol. 39.

[Str66] Strauch, R., 1966. "Negative Dynamic Programming," Ann. Math. Statist., Vol. 37, pp. 871-890.

[Str75] Striebel, 1975. Optimal Control of Discrete Time Stochastic Systems, Springer-Verlag, Berlin and New York.

[SuB98] Sutton, R. S., and Barto, A. G., 1998. Reinforcement Learning, MIT Press, Cambridge, MA.

[Sze98a] Szepesvari, C., 1998. Static and Dynamic Aspects of Optimal Sequential Decision Making, Ph.D. Thesis, Bolyai Institute of Mathematics, Hungary.

[Sze98b] Szepesvari, C., 1998. "Non-Markovian Policies in Sequential Decision Problems," Acta Cybernetica, Vol. 13, pp. 305-318.

[Sze10] Szepesvari, C., 2010. Algorithms for Reinforcement Learning, Morgan and Claypool Publishers, San Franscisco, CA.

[TBA86] Tsitsiklis, J. N., Bertsekas, D. P., and Athans, M., 1986. "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," IEEE Trans. Aut. Control, Vol. AC-31, pp. 803-812.

[ThS10a] Thiery, C., and Scherrer, B., 2010. "Least-Squares λ-Policy Iteration:

Bias-Variance Trade-off in Control Problems," in ICML'10: Proc. of the 27th Annual International Conf. on Machine Learning.

[ThS10b] Thiery, C., and Scherrer, B., 2010. "Performance Bound for Approximate Optimistic Policy Iteration," Technical Report, INRIA, France.

[Tsi94] Tsitsiklis, J. N., 1994. "Asynchronous Stochastic Approximation and Q-Learning," Machine Learning, Vol. 16, pp. 185-202.

[VVL13] Vrabie, V., Vamvoudakis, K. G., and Lewis, F. L., 2013. Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles, The Institution of Engineering and Technology, London.

[VeP87] Verdu, S., and Poor, H. V., 1987. "Abstract Dynamic Programming Models under Commutativity Conditions," SIAM J. on Control and Opt., Vol. 25, pp. 990-1006.

[Vei66] Veinott, A. F., 1966. "On Finding Optimal Policies in Discrete Dynamic Programming with no Discounting," Ann. Math. Stat., Vol. 37, pp. 1284-1294.

[Whi80] Whittle, P., 1980. "Stability and Characterization Conditions in Negative Programming," Journal of Applied Probability, Vol. 17, pp. 635-645.

[Whi90] Whittle, P., 1990. Risk-Sensitive Optimal Control, Wiley, Chichester.

[WiB93] Williams, R. J., and Baird, L. C., 1993. "Analysis of Some Incremental Variants of Policy Iteration: First Steps Toward Understanding Actor-Critic Learning Systems," Report NU-CCS-93-11, College of Computer Science, Northeastern University, Boston, MA.

[YuB10] Yu, H., and Bertsekas, D. P., 2010. "Error Bounds for Approximations from Projected Linear Equations," Math. of OR, Vol. 35, pp. 306-329.

[YuB11a] Yu, H., and Bertsekas, D. P., 2011. "Q-Learning and Policy Iteration Algorithms for Stochastic Shortest Path Problems," Lab. for Info. and Decision Systems Report LIDS-P-2871, MIT; to appear in Annals of OR; DOI: 10.1007/s10479-012-1128-z.

[YuB11b] Yu, H., and Bertsekas, D. P., 2011. "On Boundedness of Q-Learning Iterates for Stochastic Shortest Path Problems," Lab. for Info. and Decision Systems Report LIDS-P-2859, MIT; to appear in Math. of OR.

[YuB12] Yu, H., and Bertsekas, D. P., 2012. "Weighted Bellman Equations and their Applications in Dynamic Programming," Lab. for Info. and Decision Systems Report LIDS-P-2876, MIT.

[Yu11] Yu, H., 2011. "Stochastic Shortest Path Games and Q-Learning," Lab. for Info. and Decision Systems Report LIDS-P-2875, MIT.

[Yu12] Yu, H., 2012. "Least Squares Temporal Difference Methods: An Analysis Under General Conditions," SIAM J. Control and Opt., Vol. 50, pp. 3310-3343.

[Zac64] Zachrisson, L. E., 1964. "Markov Games," in Advances in Game Theory, by M. Dresher, L. S. Shapley, and A. W. Tucker, (eds.), Princeton Univ. Press, Princeton, N. J., pp. 211-253.

INDEX