

Reading Notes

Abstract Dynamic Programming¹

YUCHAO LI²

October 14, 2019

¹2nd Edition.

²<https://yuchaotaigu.github.io>

Oh, fantastic! Fantastic!
(Clive Tyldesley, on 2002 UCL final Zinedine
Zidane's winning volley.)

Contents

1	Introduction	3
1.1	Structure of Dynamic Programming	3
1.2	Abstract Dynamic Programming Models	5
1.2.1	Problem Formulation	5
1.2.2	Monotonicity and Contraction Properties	5
1.2.3	Some Examples	5
1.3	Organization of the Book	7
1.4	Notes, Sources, and Exercises	7
2	Contractive Models	11
2.1	Bellman's Equation and Optimality Conditions	11
2.2	Limited Lookahead Policies	20
2.3	Value Iteration	23
2.4	Policy Iteration	24
2.4.1	Approximate Policy Iteration	26
2.4.2	Approximate Policy Iteration Where Policies Converge	27
2.5	Optimistic Policy Iteration and λ -Policy Iteration	28
2.5.1	Convergence of Optimistic Policy Iteration	30
2.5.2	Approximate Optimistic Policy Iteration	33
3	Semicontractive Models	47
3.1	Pathologies of Noncontractive DP Models	47
3.1.1	Deterministic Shortest Path Problems	47
3.1.2	Stochastic Shortest Path Problems	47
3.1.3	The Blackmailer's Dilemma	47
3.1.4	Linear-Quadratic Problems	47
3.1.5	An Intuitive View of Semicontractive Analysis	48
3.2	Semicontractive Models and Regular Policies	49
3.2.1	S -Regular Policies	49
3.2.2	Restricted Optimization over S -Regular Policies	49

Appendices

Appendix A Notation and Mathematical Conventions	53
A.1 Set notion and conventions	53
A.2 Functions	53
Appendix B Contraction Mappings	55
B.1 Contraction mapping fixed point theorem	55
B.2 Weighted sup-norm contractions	55

Preface

1

Introduction

1.1 Structure of Dynamic Programming

P. 3

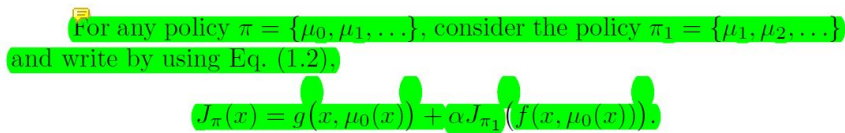


Figure 1.1: P. 3 (1).

To prove this, note that by definition,

$$\begin{aligned} J_\pi(x_0) &= \limsup_{N \rightarrow \infty} \sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k)) \\ &= \lim_{N \rightarrow \infty} \sup_{K \geq N} \sum_{k=0}^{K-1} \alpha^k g(x_k, \mu_k(x_k)) \\ &= \lim_{N \rightarrow \infty} \sup_{K \geq N} \sum_{k=0}^{K-1} a_k \\ &= \lim_{N \rightarrow \infty} \sup_{K \geq N} z_K \end{aligned} \tag{1.1}$$

where $a_k = \alpha^k g(x_k, \mu_k(x_k))$ and $z_K = \sum_{k=0}^{K-1} a_k$. Similarly, for J_{π_1} , we have

$$\begin{aligned}
J_{\pi_1}(x_1) &= \limsup_{N \rightarrow \infty} \sum_{k=1}^N \alpha^{k-1} g(x_k, \mu_k(x_k)) \\
&= \limsup_{N \rightarrow \infty} \sum_{k=1}^{N-1} \alpha^{k-1} g(x_k, \mu_k(x_k)) \\
&= \lim_{N \rightarrow \infty} \sup_{K \geq N} \sum_{k=1}^{K-1} \alpha^{k-1} g(x_k, \mu_k(x_k)) \\
&= \lim_{N \rightarrow \infty} \sup_{K \geq N} \sum_{k=1}^{K-1} b_k \\
&= \lim_{N \rightarrow \infty} \sup_{K \geq N} w_K
\end{aligned} \tag{1.2}$$

where $b_k = \alpha^{k-1} g(x_k, \mu_k(x_k))$ and $w_K = \sum_{k=1}^{K-1} b_k$. Since for any given K and x_0 , under the deterministic dynamics $x_{k+1} = f(x_k, u_k)$, it holds that $z_K = a_k + \alpha w_K$. Then consider the sequence $\{z_K\}$ and $\{w_K\}$, their limit superiors have the relation given here.

P. 3

We have for all $x \in X$

$$\begin{aligned}
J^*(x) &= \inf_{\pi = \{\mu_0, \pi_1\} \in \Pi} \left\{ g(x, \mu_0(x)) + \alpha J_{\pi_1}(f(x, \mu_0(x))) \right\} \\
&\stackrel{\text{min}}{=} \inf_{\mu_0 \in \mathcal{M}} \left\{ g(x, \mu_0(x)) + \alpha \inf_{\pi_1 \in \Pi} J_{\pi_1}(f(x, \mu_0(x))) \right\} \\
&= \inf_{\mu_0 \in \mathcal{M}} \left\{ g(x, \mu_0(x)) + \alpha J^*(f(x, \mu_0(x))) \right\}.
\end{aligned}$$

Figure 1.2: P. 3 (2).

This equality holds due to the principle of optimality.

P. 4

Defining

$$(T_\mu J)(x) = H(x, \mu(x), J), \quad x \in X,$$

and

$$(TJ)(x) = \inf_{u \in U(x)} H(x, u, J) \stackrel{\text{min}}{=} \inf_{\mu \in \mathcal{M}} (T_\mu J)(x), \quad x \in X,$$

Figure 1.3: P. 4 (1).

This equality holds due to the definition of \mathcal{M} .

P. 4

$$J_{\pi}(x) = \limsup_{N \rightarrow \infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}} \bar{J})(x), \quad x \in X. \quad (1.7)$$

Figure 1.4: P. 4 (2).

The definition here is exactly the same as the definition given by Eq. (1.2), due to the definition of the composition of mappings.

1.2 Abstract Dynamic Programming Models

1.2.1 Problem Formulation

None.

1.2.2 Monotonicity and Contraction Properties

None.

1.2.3 Some Examples

P. 11

is defined as a (possibly countably infinite) sum, since the disturbances w_k , $k = 0, 1, \dots$, take values in a countable set. Indeed, the reader may verify that all the subsequent mathematical expressions that involve an expected value can be written as summations over a finite or a countable set, so they make sense without resort to measure-theoretic integration concepts. †

Figure 1.5: P. 11 (1).

Here we have applied the Theorem that finite Cartesian product of countable sets have countable elements.

P. 11

$$(T_{\mu_0} \cdots T_{\mu_{N-1}} \bar{J})(x_0) \stackrel{E}{w_k}_{k=0,1,\dots} \left\{ \sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right\},$$

Figure 1.6: P. 11 (2).

This equality can be justified by double expectation. That is,

$$\begin{aligned}
J_{\pi,N}(x_0) &= E\left\{\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k) | x_0\right\} \\
&= E\left\{\alpha \sum_{k=1}^{N-1} \alpha^{k-1} g(x_k, \mu_k(x_k), w_k) + g(x_0, \mu_0(x_0), w_0) | x_0\right\} \\
&= E\left\{\alpha E\left\{\sum_{k=1}^{N-1} \alpha^{k-1} g(x_k, \mu_k(x_k), w_k) | x_1\right\} + g(x_0, \mu_0(x_0), w_0) | x_0\right\} \\
&= E\left\{\alpha E\left\{\dots \alpha E\left\{g(x_{N-1}, \mu_{N-1}(x_{N-1}), w_{N-1}) | x_{N-1}\right\} + \dots \right. \right. \\
&\quad \left. \left. + g(x_1, \mu_1(x_1), w_1) | x_1\right\} + g(x_0, \mu_0(x_0), w_0) | x_0\right\}.
\end{aligned}$$

Then it can be seen from the last line that the equality holds.

P. 26

By contrast, the projected equation mapping $\Pi_\xi T$ need not be monotone, because the components of Π_ξ need not be nonnegative. Moreover while the projection Π_ξ is nonexpansive with respect to the projection norm $\|\cdot\|_\xi$, it need not be nonexpansive with respect to the sup-norm. As a result the projected equation mapping $\Pi_\xi T$ need not be a sup-norm contraction. These facts play a significant role in approximate DP methodology.

Figure 1.7: P. 26.

Regarding the first comment, refer to Eq. (6.36), Section 6.3.2, P. 428 [Ber12a], or Eq. (6.77), Section 6.8, P. 355, [BeT96] for details. In short, the iteration in matrix form is given by

$$\Phi r^{k+1} = \Phi(\Phi' \Xi \Phi)^{-1} \Phi' \Xi T \Phi r^k,$$

where $\Xi \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal elements are ξ_i 's. Therefore, the projection matrix Π_ξ is

$$\Pi_\xi = \Phi(\Phi' \Xi \Phi)^{-1} \Phi' \Xi.$$

Regarding the second, the definition is given as

$$\|\Pi_\xi J\|_\xi \leq \|J\|_\xi.$$

Refer to Section 6.8, P. 355, [BeT96].

P. 28

The preceding formulas show that $T^{(\lambda)}$ and $P^{(c)}$ are closely related, and that iterating with $T^{(\lambda)}$ is “faster” than iterating with $P^{(c)}$, since the eigenvalues of A are within the unit circle, so that T is a contraction. In addition, methods such as TD(λ), LSTD(λ), LSPE(λ), and their projected versions, which are based on $T^{(\lambda)}$, can be adapted to be used with $P^{(c)}$.

Figure 1.8: P. 28 (1).

Refer to the note on Exercise 1.2 (b), P. 35 [Abstract DP] 2nd Edition for details.

P. 28

The mapping $T^{(\lambda)}$ is obtained for $w_{i\ell} = (1 - \lambda)\lambda^{\ell-1}$, independently of the state i . A more general version, where λ depends on the state i , is obtained for $w_{i\ell} = (1 - \lambda_i)\lambda_i^{\ell-1}$. The solution of Eqs. (1.24) and (1.25) by simulation-based methods is discussed in the paper [YuB12]; see also Exercise 1.3.

Figure 1.9: P. 28 (2).

Here, ‘more general’ is compared to the case where $\omega_{i\ell} = (1 - \lambda)\lambda^\ell$. The part highlighted above is the most general setting, and here $\omega_{i\ell} = (1 - \lambda_i)\lambda_i^\ell$ is one of the possible forms of $(\omega_{1\ell}, \omega_{1\ell}, \dots)$, which is a probability distribution.

1.3 Organization of the Book

None.

1.4 Notes, Sources, and Exercises**P. 35**

Thus $T^{(\lambda)}J$ is obtained by extrapolation along the line segment $P^{(c)}J - J$, as illustrated in Fig. 1.4.1. Note that since T is a contraction mapping, $T^{(\lambda)}J$ is closer to J^* than $P^{(c)}J$.

Figure 1.10: P. 35.

To see this, due to Eq. (1.28), we have

$$T^{(\lambda)}J - J^* = TP^{(c)}J - J^* = TP^{(c)}J - TJ^*.$$

Therefore, we have

$$\|T^{(\lambda)}J - J^*\| = \|TP^{(c)}J - TJ^*\| \leq \alpha \|P^{(c)}J - J^*\|.$$

P. 38

Consider a set of mappings $T_\mu : \mathcal{B}(X) \mapsto \mathcal{B}(X)$, $\mu \in \mathcal{M}$, satisfying the contraction Assumption 1.2.2. Consider also the mappings $(T_\mu^{(w)}) : \mathcal{B}(X) \mapsto \mathcal{B}(X)$ defined by

$$(T_\mu^{(w)}J)(x) = \sum_{\ell=1}^{\infty} w_\ell(x) (T_\mu^\ell J)(x), \quad x \in X, J \in \mathcal{B}(X),$$

Figure 1.11: P. 38 (1).

Refer to the note in P. 64, [Abstract DP] 2nd Edition, for further details.

P. 38

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

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

Figure 1.12: P. 38 (2).

We denote the right-hand side of the equation as $|\sum_{\ell=1}^{\infty} (a_\ell - b_\ell)|$. Due to continuity of $|\cdot|$, we have

$$\lim_{n \rightarrow \infty} \left| \sum_{\ell=1}^n (a_\ell - b_\ell) \right| = \left| \sum_{\ell=1}^{\infty} (a_\ell - b_\ell) \right|.$$

Since $\forall n$, it holds that

$$\left| \sum_{\ell=1}^n (a_\ell - b_\ell) \right| \leq \sum_{\ell=1}^n |a_\ell - b_\ell|,$$

which is due to triangular inequality, then taking limits on both sides and we get an inequality. The inequality here is obtained based on the triangular

inequality and the fact that

$$\frac{|(T_\mu^\ell J)(x) - (T_\mu^\ell J')(x)|}{v(x)} \leq \|T_\mu^\ell J - T_\mu^\ell J'\|.$$

2

Contractive Models

2.1 Bellman's Equation and Optimality Conditions

P. 40

We denote by $\mathcal{R}(X)$ the set of real-valued functions $J : X \mapsto \mathfrak{R}$. We have a mapping $H : X \times U \times \mathcal{R}(X) \mapsto \mathfrak{R}$ and for each policy $\mu \in \mathcal{M}$, we consider the mapping $T_\mu : \mathcal{R}(X) \mapsto \mathcal{R}(X)$ defined by

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

Figure 2.1: P. 40 (1).

Note that H only needs to be defined on (x, u, J) where $u \in U(x)$. For (x, u', J) where $u' \notin U(x)$, $H(x, u', J)$ can be left undefined.

P. 40

[We will use frequently the second equality above, which holds because \mathcal{M} can be viewed as the Cartesian product $\prod_{x \in X} U(x)$.] We want to find a function $J^* \in \mathcal{R}(X)$ such that

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

i.e., to find a fixed point of T within $\mathcal{R}(X)$. We also want to obtain a policy $\mu^* \in \mathcal{M}$ such that $T_{\mu^*} J^* = T J^*$.

Figure 2.2: P. 40 (2).

J^* is defined as the fixed point of T within $\mathcal{R}(X)$.

P. 41

Note that the monotonicity assumption implies the following properties, for all $J, J' \in \mathcal{R}(X)$ and $k = 0, 1, \dots$, which we will use extensively:

$$\begin{aligned} J \leq J' &\Rightarrow T^k J \leq T^k J', & T_\mu^k J \leq T_\mu^k J', & \forall \mu \in \mathcal{M}, \\ J \leq TJ &\Rightarrow T^k J \leq T^{k+1} J, & T_\mu^k J \leq T_\mu^{k+1} J, & \forall \mu \in \mathcal{M}. \end{aligned}$$

Figure 2.3: P. 41 (1).

Regarding the first comment, given the assumption $H : X \times U \times \mathcal{R}(X) \rightarrow \mathbb{R}$, we have that $\forall J \in \mathcal{R}(X)$, it holds that $T_\mu J \in \mathcal{R}(X) \forall \mu \in \mathcal{M}$. However, it does not imply $TJ \in \mathcal{R}(X)$. Therefore, to have the k -folds well-defined, e.g. $T(TJ)$ well-defined, one need to ensure first $TJ \in \mathcal{R}(X)$ so that $T(TJ)$ can be defined.

Regarding the second comment, as noted above, $H : X \times U \times \mathcal{R}(X) \rightarrow \mathbb{R}$ does not imply $TJ \in \mathcal{R}(X) \forall J \in \mathcal{R}(X)$. However, given $J \in \mathcal{R}(X)$, if in addition, we have $J \leq TJ$, then $T_\mu J(x)$ is lower bounded by $J(x) \forall \mu \in \mathcal{M}$, $\forall x \in X$, so $TJ(x) = \inf_{\mu \in \mathcal{M}} T_\mu J(x) \in \mathbb{R}$, which means $TJ \in \mathcal{R}(X)$ for the given J .

P. 41

Assumption 2.1.2: (Contraction) For all $J \in \mathcal{B}(X)$ and $\mu \in \mathcal{M}$, the functions $T_\mu J$ and TJ belong to $\mathcal{B}(X)$. Furthermore, for some $\alpha \in (0, 1)$, we have

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

Figure 2.4: P. 41 (2).

Refer to Prop. B.5, P. 333, [Abstract DP] 2nd Edition for an example.

P. 42

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

$$\lim_{k \rightarrow \infty} \|J^* - T^k J\| = 0, \quad \lim_{k \rightarrow \infty} \|J_\mu - T_\mu^k J\| = 0.$$

Figure 2.5: P. 42 (1).

According to Prop. B.1, P. 327, [Abstract DP] 2nd Edition, J^* and J_μ are fixed points of T and T_μ respectively and the convergence is defined in terms of the weighted sup-norm. Here under the assumption that $TJ \in \mathcal{B}(X)$, $T_\mu J \in \mathcal{B}(X)$ holds $\forall J \in \mathcal{B}(X)$, $\forall \mu \in \mathcal{M}$ [this is part of the Assumption 2.1.1], we would like to show that convergence in norm implies point-wise convergence. To see this, note that by definition, we have

$$\|T^k J - J^*\| = \sup_{x \in X} \frac{|T^k J(x) - J^*(x)|}{v(x)},$$

therefore, $\forall x \in X$, it holds that

$$|T^k J(x) - J^*(x)| \leq v(x) \|T^k J - J^*\|.$$

Since $T^k J \rightarrow J^*$ in norm, namely $\lim_{k \rightarrow \infty} \|T^k J - J^*\| = 0$, then by first taking limit inferiors on both sides of above equation, and then taking limit superiors on both sides of above equation, we get the point-wise convergence. Replace T and J^* with T_μ and J_μ respectively, the exact same arguments can be applied to prove the point-wise convergence $(T_\mu J)(x) \rightarrow J_\mu(x)$. Note that in our proof for the convergence in norm implying convergence point-wise, we only requires that $TJ \in \mathcal{B}(X)$, $T_\mu J \in \mathcal{B}(X)$ holds $\forall J \in \mathcal{B}(X)$, $\forall \mu \in \mathcal{M}$; the contraction property of T and T_μ is not used.

Due to above result, we have

$$\limsup_{k \rightarrow \infty} T^k J(x) = \lim_{k \rightarrow \infty} T^k J(x) = J^*(x).$$

The same result goes for T_μ .

P. 42

To show part (d), we use the triangle inequality to write for every k ,

$$\|T^k J - J\| \leq \sum_{\ell=1}^k \|T^\ell J - T^{\ell-1} J\| \leq \sum_{\ell=1}^k \alpha^{\ell-1} \|TJ - J\|.$$

Taking the limit as $k \rightarrow \infty$ and using part (b), the left-hand side inequality follows. The right-hand side inequality follows from the left-hand side and the contraction property of T . The proof of part (e) is similar to part (d) [indeed it is the special case of part (d) where T is equal to T_μ , i.e., when $U(x) = \{\mu(x)\}$ for all $x \in X$]. **Q.E.D.**

Figure 2.6: P. 42 (2).

The following four lemmas are needed in the follow up discussion. Refer to [Hand Note 4] for proofs.

Lemma (1, P. 42). *Given two extended real-valued sequences $\{a_n\}$ and $\{b_n\}$ where $a_n, b_n \in \mathbb{R}^*$. Assume that both $\lim_{n \rightarrow \infty} a_n$ and $\lim_{n \rightarrow \infty} b_n$ exist in \mathbb{R}^* . Prove that*

$$a_n \leq b_n \implies \lim_{n \rightarrow \infty} a_n \leq \lim_{n \rightarrow \infty} b_n.$$

Lemma (2, P. 42). *Given $\{a_n\}$ with $a_n \in \mathbb{R}$ and $a = \lim_{n \rightarrow \infty} a_n \in \mathbb{R}$, and $\{b_n\}$ with $b_n \in \mathbb{R}^*$ and $b = \lim_{n \rightarrow \infty} b_n \in \mathbb{R}^*$. Prove that*

$$\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n.$$

Lemma (3, P. 42). *Given $\{a_n\} \in \mathbb{R}^*$ which is monotone, prove that $\{a_n\}$ is convergent in \mathbb{R}^* .*

Lemma (4, P. 42). *Given two real sequences $\{a_n\}$ and $\{b_n\}$, and assume that $\limsup_{n \rightarrow \infty} a_n \in \mathbb{R}$, show that*

$$\limsup_{n \rightarrow \infty} (a_n + b_n) \leq \limsup_{n \rightarrow \infty} a_n + \limsup_{n \rightarrow \infty} b_n.$$

There are two approaches for proof of the highlighted part.

1. For any k , we have

$$\begin{aligned} \|J^* - J\| &\leq \|J^* - T^k J\| + \|T^k J - J\| \\ &\leq \|J^* - T^k J\| + \sum_{\ell=1}^k \alpha^{\ell-1} \|TJ - J\|. \end{aligned}$$

Taking limit on k , we get the desired result. Here we can directly take limit on k since $\lim_{k \rightarrow \infty} \|J^* - T^k J\|$ exists and the sequence $\{\sum_{\ell=1}^k \alpha^{\ell-1} \|TJ - J\|\}_{k=1}^\infty$ is monotonically nondecreasing.

2. Since $T^k J \rightarrow J^*$ in the sense that $\lim_{k \rightarrow \infty} \|T^k J - J^*\| = 0$, due to continuity of $\|\cdot\|$, $\|T^k J - J\| \rightarrow \|J^* - J\|$. [A closer look of this arguments is given below.]

Theorem (P. 42). $a_n \rightarrow a$ in the sense that $\lim_{n \rightarrow \infty} \|a_n - a\| = 0$, then $\lim_{n \rightarrow \infty} \|a_n\| = \|\lim_{n \rightarrow \infty} a_n\| = \|a\|$.

Proof. Since we have $\|a_n\| = \|a_n - a + a\| \leq \|a_n - a\| + \|a\|$, then $\sup_{k \geq n} \|a_k\| \leq \sup_{k \geq n} (\|a_k - a\| + \|a\|)$, then by [Lemma 3 P. 42], both are convergent in \mathbb{R}^* and by [Lemma 1 P. 42],

$$\limsup_{n \rightarrow \infty} \|a_n\| \leq \limsup_{n \rightarrow \infty} (\|a_n - a\| + \|a\|).$$

Since $\limsup_{n \rightarrow \infty} \|a\| = \|a\| \in \mathbb{R}$, by [Lemma 4 P. 42],

$$\limsup_{n \rightarrow \infty} (\|a_n - a\| + \|a\|) \leq \limsup_{n \rightarrow \infty} \|a_n - a\| + \|a\| = \|a\|.$$

On the other hand, we have

$$\liminf_{n \rightarrow \infty} \|a_n\| = \liminf_{n \rightarrow \infty} \|a + a_n - a\| \geq \liminf_{n \rightarrow \infty} (\|a\| - \|a_n - a\|) = \|a\|,$$

which concludes the proof. \square

P. 42

Part (c) of the preceding proposition shows that there exists a $\mu \in \mathcal{M}$ such that $J_\mu = J^*$ if and only if the minimum of $H(x, u, J^*)$ over $U(x)$ is attained for all $x \in X$. Of course the minimum is attained if $U(x)$ is finite for every x , but otherwise this is not guaranteed in the absence of additional assumptions. Part (d) provides a useful error bound: we can evaluate the proximity of any function $J \in \mathcal{B}(X)$ to the fixed point J^* by applying T to J and computing $\|TJ - J\|$. The left-hand side inequality of part (e) (with $J = J^*$) shows that for every $\epsilon > 0$, there exists a $\mu_\epsilon \in \mathcal{M}$ such that $\|J_{\mu_\epsilon} - J^*\| \leq \epsilon$, which may be obtained by letting $\mu_\epsilon(x)$ minimize $H(x, u, J^*)$ over $U(x)$ within an error of $(1 - \alpha)\epsilon v(x)$, for all $x \in X$.

Figure 2.7: P. 42 (3).

To see this, note that for a given x , we have

$$J^*(x) = TJ^*(x) = \left(\inf_{\mu \in \mathcal{M}} T_\mu J^* \right)(x) = \inf_{u \in U(x)} H(x, u, J^*).$$

If the infimum is attained, we can define

$$\mu_\epsilon(x) = \arg \min_{u \in U(x)} H(x, u, J^*);$$

otherwise, due to the definition of infimum, $(\forall \varepsilon > 0)(\exists u_\varepsilon \in U(x))(H(x, u_\varepsilon, J^*) < \inf_{u \in U(x)} H(x, u, J^*) + \varepsilon = J^*(x) + \varepsilon)$. Then we can define

$$\mu_\varepsilon(x) = u_\varepsilon.$$

Note that the above construction relies on \mathcal{M} being the Cartesian product of feasible control sets $U(x)$.

P. 43

Proposition 2.1.2: Let the monotonicity and contraction Assumptions 2.1.1 and 2.1.2 hold. Then

$$J^*(x) = \inf_{\mu \in \mathcal{M}} J_\mu(x), \quad \forall x \in X.$$

Furthermore, for every $\varepsilon > 0$, there exists $\mu_\varepsilon \in \mathcal{M}$ such that

$$J^*(x) \leq J_{\mu_\varepsilon}(x) \leq J^*(x) + \varepsilon, \quad \forall x \in X. \quad (2.1)$$

Figure 2.8: P. 43 (1).

The left-hand side is the fixed point of T ; while the right-hand side is the point-wise infimum of $J_\mu(x)$ over \mathcal{M} .

P. 43

Note that **without monotonicity**, we may have $\inf_{\mu \in \mathcal{M}} J_\mu(x) < J^*(x)$ for some x . This is illustrated by the following example.

Figure 2.9: P. 43 (2).

Here it means that without monotonicity assumption but with contraction assumption.

P. 44

$$J_\pi(x) = \limsup_{k \rightarrow \infty} (T_{\mu_0} \cdots T_{\mu_k} J)(x), \quad \forall x \in X,$$

Figure 2.10: P. 44 (1).

Under Assumption 2.1.2 (Assumption 2.1.1 is irrelevant here), given $\bar{J} \in \mathcal{B}(X)$, it is still possible that $J_\pi \notin \mathcal{B}(X)$. In fact, J_π can even take value $\pm\infty$ at some x , which is to say it's possible that $J_\pi \notin \mathcal{R}(X)$. One example could be consider the case where the state space is a singleton and control space is infinite with accumulative cost. For μ_0 , we have one stage cost g_0 ; then the stage cost of μ_k is defined as g_0/α^k where α is the discount factor for all T_{μ_k} . Then it can be shown that $J_\pi = g_0 \cdot (+\infty)$. This indicates that given initial $\bar{J} \in \mathcal{B}(X)$, it could be that $J_\pi \notin \mathcal{R}(X)$. In particular, the proof technique used in Prop. B.1, P. 327, [Abstract DP] 2nd Edition, does not work in this case. To see this, note the sequence generated according to the definition is

$$\bar{J}, T_{\mu_0}\bar{J}, T_{\mu_0}T_{\mu_1}\bar{J}, \dots, T_{\mu_0}T_{\mu_1}\cdots T_{\mu_k}\bar{J}, T_{\mu_0}T_{\mu_1}\cdots T_{\mu_k}T_{\mu_{k+1}}\bar{J}, \dots \quad (2.1)$$

Therefore, the normed difference of adjacent terms is bounded by

$$\|T_{\mu_0}\bar{J} - \bar{J}\|, \alpha\|T_{\mu_1}\bar{J} - \bar{J}\|, \dots, \alpha^k\|T_{\mu_{k+1}}\bar{J} - \bar{J}\|, \dots, \quad (2.2)$$

where the k th term in (2.2) is the bound of the difference of the k th and $k+1$ th terms in (2.1). Here we have applied the result that if $J, J' \in \mathcal{B}(X)$, then $J - J' \in \mathcal{B}(X)$ (so that the differences of any adjacent terms in (2.1) is in $\mathcal{B}(X)$ and therefore can be plugged into $\|\cdot\|$). Note that although due to contraction assumption, we have a geometric term α^k ; however, if $\|T_{\mu_{k+1}}\bar{J} - \bar{J}\|/\|T_{\mu_k}\bar{J} - \bar{J}\| = 1/\alpha$, then the sequence would not be Cauchy, just as the example indicated.

However, under Assumptions 2.1.1 and 2.1.2, we have $J_\pi(x) > -\infty$ $\forall x \in X$ and $\forall \pi \in \Pi$. Since by Assumption 2.1.1, we have $T_{\mu_0}T_{\mu_1}\cdots T_{\mu_k}\bar{J} \geq T^{k+1}\bar{J}$. Taking limit supremum on both sides, and applying the fact that convergence in norm implies convergence point-wise where we have used the Assumption 2.1.2 [Note in P. 42], we see $J_\pi(x) \geq J^*(x) > -\infty$.

Theorem (P. 44). *Given $J_k \in \mathcal{B}(X)$, denote its point-wise limit superior as J , namely*

$$J(x) = \limsup_{k \rightarrow \infty} J_k(x),$$

then we claim the following statement:

$$J \notin \mathcal{B}(X) \implies \limsup_{k \rightarrow \infty} \|J_k\| = \infty. \quad (2.3)$$

Proof. It's equivalent to prove the following statement:

$$\limsup_{k \rightarrow \infty} \|J_k\| < \infty \implies J \in \mathcal{B}(X). \quad (2.4)$$

Since $\limsup_{k \rightarrow \infty} \|J_k\| < \infty$, denote $A = \limsup_{k \rightarrow \infty} \|J_k\|$, then $\forall x \in X$, it holds that

$$\frac{|J_k(x)|}{v(x)} \leq \|J_k\| \implies \limsup_{k \rightarrow \infty} \frac{|J_k(x)|}{v(x)} \leq \limsup_{k \rightarrow \infty} \|J_k\| = A.$$

Since $\forall x \in X$ and k , we also have $-|J_k(x)| \leq J_k(x) \leq |J_k(x)|$, then it holds that

$$\liminf_{k \rightarrow \infty} (-|J_k(x)|) = -\limsup_{k \rightarrow \infty} |J_k(x)| \leq \limsup_{k \rightarrow \infty} J_k(x) \leq \limsup_{k \rightarrow \infty} |J_k(x)|.$$

Therefore, we have $\forall x \in X$,

$$\frac{|J(x)|}{v(x)} = \frac{|\limsup_{k \rightarrow \infty} J_k(x)|}{v(x)} \leq \limsup_{k \rightarrow \infty} \frac{|J_k(x)|}{v(x)} \leq A.$$

Then we have

$$\sup_{x \in X} \frac{|J(x)|}{v(x)} \leq A.$$

□

With the [Theorem P. 44], we know that if $J_\pi \notin \mathcal{B}(X)$, then

$$\limsup_{k \rightarrow \infty} \|T_{\mu_0} \cdots T_{\mu_k} \bar{J}\| = \infty.$$

P. 44

Note that under the contraction Assumption 2.1.2, the choice of \bar{J} in the definition of J_π does not matter, since for any two $J, J' \in \mathcal{B}(X)$, we have

$$\|T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k} J - T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k} J'\| \leq \alpha^{k+1} \|J - J'\|,$$

so the value of $J_\pi(x)$ is independent of \bar{J} . Since by Prop. 2.1.1(b), $J_\mu(x) = \lim_{k \rightarrow \infty} (T_\mu^k J)(x)$ for all $\mu \in \mathcal{M}$, $J \in \mathcal{B}(X)$, and $x \in X$, in the DP context we recognize J_μ as the cost function of the stationary policy $\{\mu, \mu, \dots\}$.

Figure 2.11: P. 44 (2).

By definition of $\|\cdot\|$, we have

$$\sup_{x \in X} \frac{|(T_{\mu_0} \cdots T_{\mu_k} J)(x) - (T_{\mu_0} \cdots T_{\mu_k} J')(x)|}{v(x)} \leq \alpha^{k+1} \|J - J'\|,$$

then $\forall x \in X$ and $k \in \mathbb{N}$, we have

$$|a_k(x) - a'_k(x)| \leq v(x) \alpha^{k+1} \|J - J'\|$$

where $a_k(x) = (T_{\mu_0} \cdots T_{\mu_k} J)(x)$ and $a'_k(x) = (T_{\mu_0} \cdots T_{\mu_k} J')(x)$. Then we have $\limsup_{k \rightarrow \infty} (a_k(x) - a'_k(x)) = \limsup_{k \rightarrow \infty} (a'_k(x) - a_k(x)) = 0$. Since we have

$$\begin{aligned} \limsup_{k \rightarrow \infty} a_k(x) &= \limsup_{k \rightarrow \infty} (a_k(x) - a'_k(x) + a'_k(x)) \\ &\leq \limsup_{k \rightarrow \infty} (a_k(x) - a'_k(x)) + \limsup_{k \rightarrow \infty} a'_k(x), \end{aligned}$$

where the inequality is due to [Lemma 4 P. 42]. Then it holds that

$$\begin{aligned}
 \limsup_{k \rightarrow \infty} a_k(x) &\leq \limsup_{k \rightarrow \infty} (a_k(x) - a'_k(x)) + \limsup_{k \rightarrow \infty} a'_k(x) \\
 &\leq \limsup_{k \rightarrow \infty} v(x) \alpha^{k+1} \|J - J'\| + \limsup_{k \rightarrow \infty} a'_k(x) \\
 &= \lim_{k \rightarrow \infty} v(x) \alpha^{k+1} \|J - J'\| + \limsup_{k \rightarrow \infty} a'_k(x) \\
 &= \limsup_{k \rightarrow \infty} a'_k(x).
 \end{aligned}$$

Note that here we avoid to use $\limsup_{k \rightarrow \infty} a_k(x) - \limsup_{k \rightarrow \infty} a'_k(x) \leq \limsup_{k \rightarrow \infty} (a_k(x) - a'_k(x))$ in order to avoid possible $\infty - \infty$. Swap the order of $a_k(x)$ and $a'_k(x)$, we get

$$\limsup_{k \rightarrow \infty} a'_k(x) \leq \limsup_{k \rightarrow \infty} a_k(x),$$

which concludes the proof.

P. 44

$$J_\pi(x) = \limsup_{k \rightarrow \infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k} \bar{J})(x) \geq \lim_{k \rightarrow \infty} (T^{k+1} \bar{J})(x) = J^*(x)$$

Figure 2.12: P. 44 (3).

Here we have applied the fact that convergence in norm implies point-wise convergence [Theorem P. 42] and therefore, under Assumption 2.1.2, the equation here holds.

2.2 Limited Lookahead Policies

P. 50

Proposition 2.2.2: (Multistep Lookahead Error Bound) Let the contraction Assumption 2.1.2 hold. The periodic policy

$$\pi = \{\mu_0, \dots, \mu_{m-1}, \mu_0, \dots, \mu_{m-1}, \dots\}$$

generated by the method of Eq. (2.10) satisfies

$$\|J_{\pi} - J^*\| \leq \frac{2\alpha^m}{1 - \alpha^m} \|J_m - J^*\| + \frac{\epsilon}{1 - \alpha^m} + \frac{\alpha(\epsilon + 2\delta)(1 - \alpha^{m-1})}{(1 - \alpha)(1 - \alpha^m)}. \quad (2.11)$$

Figure 2.13: P. 50.

As noted in P. 44, given a nonstationary policy π , under Assumption 2.1.2 (Assumption 2.1.1 is irrelevant), it could be that $J_{\pi} \notin \mathcal{R}(X)$. However, for a periodic policy π , we always have $J_{\pi} \in \mathcal{B}(X)$. To see this, define $T_{\pi_m} = T_{\mu_0} \cdots T_{\mu_{m-1}}$. By Exercise 1.1, P. 33, we have T_{π_m} contraction with modulus α^m . Denote its fixed point as $J_{\pi_m} \in \mathcal{B}(X)$. Given the initial $\bar{J} \in \mathcal{B}(X)$, the sequence generated by applying k -folds of T_{π_m} is

$$\bar{J}, T_{\pi_m} \bar{J}, \dots, T_{\pi_m}^k \bar{J}, \dots \quad (2.5)$$

which is Cauchy and converges in norm to $J_{\pi_m} \in \mathcal{B}(X)$.

On the other hand, according to the definition given in P. 44,

$$J_{\pi} = \limsup_{k \rightarrow \infty} T_{\mu_0} \cdots T_{\mu_k} \bar{J},$$

then the sequence generated according to the definition is given as

$$\bar{J}, T_{\mu_0} \bar{J}, T_{\mu_0} T_{\mu_1} \bar{J}, \dots, T_{\pi_m} \bar{J}, T_{\pi_m} T_{\mu_0} \bar{J}, \dots \quad (2.6)$$

We claim the sequence (2.6) is Cauchy and converges to $J_{\pi} \in \mathcal{B}(X)$, which is equal to J_{π_m} . To see this, denote

$$\Delta = \max_{i \in \{0, 1, \dots, m-2\}} \|a_i - \bar{J}\|$$

where

$$\{a_0, a_1, \dots, a_{m-2}\} = \{T_{\mu_0} \bar{J}, T_{\mu_0} T_{\mu_1} \bar{J}, \dots, T_{\mu_0} \cdots T_{\mu_{m-2}} \bar{J}\}.$$

Then $\forall \epsilon > 0$, $\exists N_1(\epsilon)$, such that $\alpha^{mi} \Delta < \epsilon/3$ holds $\forall i > N_1(\epsilon)$; besides, since sequence (2.5) Cauchy, $\exists N_2(\epsilon)$ such that $\|T_{\pi_m}^i \bar{J} - T_{\pi_m}^j \bar{J}\| < \epsilon/3$ holds

$\forall i, j > N_2(\varepsilon)$. Denote p th and q th term of the sequence (2.6) as b_p and b_q , then $\forall p, q > m \cdot \max\{N_1(\varepsilon), N_2(\varepsilon)\}$ where $\lfloor p/m \rfloor \neq 0$ and $\lfloor q/m \rfloor \neq 0$, it holds that

$$\begin{aligned} \|b_p - b_q\| &= \|T_{\pi_m}^i T_{\mu_0} \cdots T_{\mu_\ell} \bar{J} - T_{\pi_m}^j T_{\mu_0} \cdots T_{\mu_k} \bar{J}\| \\ &\leq \|T_{\pi_m}^i T_{\mu_0} \cdots T_{\mu_\ell} \bar{J} - T_{\pi_m}^i \bar{J}\| + \|T_{\pi_m}^i \bar{J} - T_{\pi_m}^j \bar{J}\| + \\ &\quad \|T_{\pi_m}^j T_{\mu_0} \cdots T_{\mu_k} \bar{J} - T_{\pi_m}^j \bar{J}\| \\ &< \alpha^{mi} \Delta + \varepsilon/3 + \alpha^{mj} \Delta \\ &< \varepsilon. \end{aligned}$$

where $i = \lfloor p/m \rfloor$, $\ell = p - mi - 1$, $j = \lfloor q/m \rfloor$, and $k = p - mj - 1$. For the cases where $\lfloor p/m \rfloor = 0$ or $\lfloor q/m \rfloor = 0$, the same holds. Therefore, we have sequence (2.6) Cauchy. By the same arguments, it can be shown that its fixed point J_π , is the same as J_{π_m} . Another way to see this is to regard sequence (2.6) as the following m subsequences:

$$\begin{aligned} &\bar{J}, T_{\pi_m} \bar{J}, \dots, T_{\pi_m}^k \bar{J}, \dots \\ &T_{\mu_0} \bar{J}, T_{\pi_m} T_{\mu_0} \bar{J}, \dots, T_{\pi_m}^k T_{\mu_0} \bar{J}, \dots \\ &\vdots \\ &T_{\mu_0} \cdots T_{\mu_{m-2}} \bar{J}, T_{\pi_m} T_{\mu_0} \cdots T_{\mu_{m-2}} \bar{J}, \dots, T_{\pi_m}^k T_{\mu_0} \cdots T_{\mu_{m-2}} \bar{J}, \dots \end{aligned}$$

all of which converge to the fixed point of T_{π_m} , then we can apply the same arguments used in P. 329 for proving Prop. B.2, we can see the sequence (2.6) converges to J_π .

However, with the given initial $\bar{J} \in \mathcal{B}(X)$, if we consider the sequence directly generated by the multistep look ahead algorithm, which could be of the form

$$\bar{J}, T_{\mu_{m-1}} \bar{J}, T_{\mu_{m-2}} T_{\mu_{m-1}} \bar{J}, \dots, T_{\pi_m} \bar{J}, T_{\mu_{m-1}} T_{\pi_m} \bar{J}, \dots \quad (2.7)$$

We claim the sequence is not convergent. To see this, note that it can be regarded as m subsequences

$$\begin{aligned} &\bar{J}, T_{\pi_m^0} \bar{J}, \dots, T_{\pi_m^0}^k \bar{J}, \dots \\ &T_{\mu_{m-1}} \bar{J}, T_{\pi_m^{m-1}} T_{\mu_{m-1}} \bar{J}, \dots, T_{\pi_m^{m-1}}^k T_{\mu_{m-1}} \bar{J}, \dots \\ &\vdots \\ &T_{\mu_1} \cdots T_{\mu_{m-1}} \bar{J}, T_{\pi_m^1} T_{\mu_1} \cdots T_{\mu_{m-1}} \bar{J}, \dots, T_{\pi_m^1}^k T_{\mu_1} \cdots T_{\mu_{m-1}} \bar{J}, \dots \end{aligned}$$

where $T_{\pi_m^l} = T_{\mu_l} \cdots T_{\mu_{m-1}} T_{\mu_0} \cdots T_{\mu_{l-1}}$ are contractions with modulus α^m . Since the contractions are different for different values of l , then those subsequences converge to different fixed points. To see this, consider one example where $T_1 J = 5 + 0.2J$ and $T_2 J = 9 + 0.2J$. Then we have $T_1 T_2 J = 9.5 + 0.04J$ and $T_2 T_1 J = 10 + 0.04J$, which have different fixed points.

P. 51

We also have using Prop. 2.1.1(e), applied in the context of the multistep mapping of Example 1.3.1,

$$\|J_\pi - J^*\| \leq \frac{1}{1 - \alpha^m} \|T_{\mu_0} \cdots T_{\mu_{m-1}} J^* - J^*\|.$$

Combining the last two relations, we obtain the desired result. **Q.E.D.**

Figure 2.14: P. 51.

The Prop. 2.1.1(e) is in fact the starting point of the proof. To understand the derivation of this bound, the proof may be read from here and proceed backwards.

P. 52

We finally note that Prop. 2.2.2 shows that as the lookahead size m increases, the corresponding bound for $\|J_\pi - J^*\|$ tends to $\epsilon + \alpha(\epsilon + 2\delta)/(1 - \alpha)$, or

$$\limsup_{m \rightarrow \infty} \|J_\pi - J^*\| \leq \frac{\epsilon + 2\alpha\delta}{1 - \alpha}.$$

Figure 2.15: P. 52.

Denote the periodic policy with period m as $\pi(m)$. The sequence here is well defined since as proved in P. 50, for every m , $J_{\pi(m)} \in \mathcal{B}(X)$, therefore, $J_{\pi(m)} - J^* \in \mathcal{B}(X)$.

In addition, due to [Theorem P. 44], we know the point-wise limit superior of $J_{\pi(m)} - J^*$, and, therefore, the point-wise limit superior of $J_{\pi(m)}$, are in the space of $\mathcal{B}(X)$. Namely, we have $J \in \mathcal{B}(X)$ where

$$J(x) = \limsup_{m \rightarrow \infty} J_{\pi(m)}(x).$$

Note that J is the point-wise limit superior of the sequence

$$J_{\pi(1)}, J_{\pi(2)}, J_{\pi(3)}, \dots$$

whose every element is a cost function of a policy but J is most likely not to be a cost function of any policy. In comparison, the definition of J_π where π is a nonstationary policy is the point-wise limit superior of the sequence

$$T_{\mu_0} \bar{J}, T_{\mu_0} T_{\mu_1} \bar{J}, T_{\mu_0} T_{\mu_1} T_{\mu_2} \bar{J}, \dots$$

whose every element is likely not to be a cost function of any policy but J_π is the cost function of π .

However, one need to note that $\|J - J^*\| \neq \limsup_{m \rightarrow \infty} \|J_{\pi(m)} - J^*\|$. One example could be $X = [0, 1]$, $f(x) = 0$, and

$$f_m(x) = \begin{cases} 1, & x \in (0, 1/m] \\ 0, & \text{o.w.} \end{cases}$$

Then we have $f_m \rightarrow f$ point-wise, but $\|f\|_\infty = 0 < \limsup_{m \rightarrow \infty} \|f_m\|_\infty = 1$.

2.3 Value Iteration

P. 54

and finally

$$\|J_k - T^k J_0\| \leq \frac{(1 - \alpha^k)\delta}{1 - \alpha}, \quad k = 0, 1, \dots \quad (2.21)$$

By taking limit as $k \rightarrow \infty$ and by using the fact $\lim_{k \rightarrow \infty} T^k J_0 = J^*$, we obtain Eq. (2.19).

Figure 2.16: P. 54.

Theorem (P. 54). *Given $\limsup_{k \rightarrow \infty} \|b_k - b\| = 0$, show that*

$$\limsup_{k \rightarrow \infty} \|a_k - b_k\| = \limsup_{k \rightarrow \infty} \|a_k - b\|.$$

Proof. By triangular inequality, we have

$$\|a_k - b_k\| \leq \|a_k - b\| + \|b - b_k\|.$$

Take limit superior on both sides. Since $\limsup_{k \rightarrow \infty} \|b_k - b\| = 0$, due to [Lemma 4 P. 42] we have

$$\limsup_{k \rightarrow \infty} \|a_k - b_k\| \leq \limsup_{k \rightarrow \infty} \|a_k - b\|.$$

Similarly, by triangular inequality, we have

$$\|a_k - b\| \leq \|a_k - b_k\| + \|b - b_k\|.$$

Taking limit superior on both sides and due to $\limsup_{k \rightarrow \infty} \|b_k - b\| = 0$, we have

$$\limsup_{k \rightarrow \infty} \|a_k - b\| \leq \limsup_{k \rightarrow \infty} \|a_k - b_k\|,$$

which concludes the proof. \square

2.4 Policy Iteration

P. 56

We assume that the minimum of $H(x, u, J_{\mu^k})$ over $u \in U(x)$ is attained for all $x \in X$, so that the improved policy μ^{k+1} is defined (we use this assumption for all the PI algorithms of the book). The following proposition establishes a basic cost improvement property, as well as finite convergence for the case where the set of policies is finite.

Figure 2.17: P. 56.

[Assumption P. 56].

P. 58

(b) Each control constraint set $U(x)$, $x = 1, \dots, n$, is a compact subset of \mathbb{R}^m .

Figure 2.18: P. 58 (1).

Theorem (P. 58). *Given $E \subset \mathbb{R}$ being compact, we have $\inf E \in E$.*

Proof. Due to the definition of $\inf E$, $\forall \varepsilon > 0$, $\exists e \in E$ such that $e \in (\inf E - \varepsilon, \inf E + \varepsilon)$; otherwise, if $\exists \varepsilon_0 > 0$ such that $(\inf E - \varepsilon_0, \inf E + \varepsilon_0) \cap E = \emptyset$, then $\inf E + \varepsilon_0$ is another lower bound, which contradicts the definition of $\inf E$. As a result, we have $\inf E$ is a limit point of E . Since E is compact hence closed, then $\inf E \in E$. \square

The compactness of $U(x)$ and continuity of $H(x, \cdot, \cdot)$ ensures that the [Assumption P. 56] is fulfilled. To see this, note that given $x \in X$ and $J \in \mathcal{R}(X)$, since $H(x, \cdot, J)$ is continuous and $U(x)$ is compact, we have that $H(x, U(x), J) \subset \mathbb{R}$ is a compact set. By [Theorem P. 58], $\inf H(x, U(x), J) \in H(x, U(x), J)$. Then $U(x) \cap H^{-1}(x, \inf H(x, U(x), J), J) \neq \emptyset$. Namely, the minimum is attained in $U(x)$.

P. 58

(c) Each function $H(x, \cdot, \cdot)$, $x = 1, \dots, n$, is continuous over $U(x) \times \mathbb{R}^n$.

Figure 2.19: P. 58 (2).

Note that the condition here is for a given x , function $H(x, \cdot, \cdot)$ is continuous on the product space $A = U(x) \times \mathbb{R}^n$. If $U(x)$ is equipped with $\|\cdot\|_2$, \mathbb{R}^n is with $\|\cdot\|$ (weighted sup-norm), then for the product space A we can equip an product metric

$$d_\infty(\cdot, \cdot) = \max\{d_{\|\cdot\|_2}(\cdot, \cdot), d_{\|\cdot\|}(\cdot, \cdot)\}$$

where $d_{\|\cdot\|_2}(\cdot, \cdot)$, $d_{\|\cdot\|}(\cdot, \cdot)$ are induced metrics on $U(x)$ and \mathbb{R}^n respectively. Namely, for $a_1, a_2 \in A$ where $a_i = (u_i, J_i)$ $i = 1, 2$, we have $d_\infty(a_1, a_2) = \max\{d_{\|\cdot\|_2}(u_1, u_2), d_{\|\cdot\|}(J_1, J_2)\}$. Then with $H(x, \cdot, \cdot)$ being continuous on the product space $A = U(x) \times \mathbb{R}^n$, we know that if $(u_k, J_k) \rightarrow (u, J)$ in the sense of d_∞ , it holds that

$$\lim_{k \rightarrow \infty} H(x, u_k, J_k) = H(x, u, J).$$

If we view

$$H(x, u_m, J_n) = s(m, n)$$

as a double sequence, then continuity of $H(x, \cdot, \cdot)$ implies that if u_m, J_n are convergent respectively, then the double sequence $s(m, n)$ has double limit (refer to [Note 1] for definitions). Namely, if $u_m \rightarrow u$ and $J_n \rightarrow J$ in the sense of their respective norms, then $s(m, n)$ is convergent and its double limit is $H(x, u, J)$. To see this, denote $a = H(x, u, J)$. $\forall \varepsilon > 0$, denote

$$V_\varepsilon = \{(u', J') | H(x, u', J') \in (a - \varepsilon, a + \varepsilon)\}.$$

Since H continuous, $V_\varepsilon \subset A$ is open and $(u, J) \in V_\varepsilon$. Then $\exists \delta > 0$ such that $B_\delta \subset V_\varepsilon$ where $B_\delta = \{(u', J') | d_\infty((u', J'), (u, J)) < \delta\}$. Since $u_m \rightarrow u$ and $J_n \rightarrow J$, for the given δ , $\exists N_1, N_2$ such that $d_{\|\cdot\|_2}(u_m, u) < \delta$ and $d_{\|\cdot\|}(J_n, J) < \delta \forall m > N_1, n > N_2$. Therefore, $\forall m, n > \max\{N_1, N_2\}$, we have

$$(u_m, J_n) \in B_\delta \implies (u_m, J_n) \in V_\varepsilon \implies H(x, u_m, J_n) \in (a - \varepsilon, a + \varepsilon),$$

which indeed implies $|s(m, n) - a| < \varepsilon$.

P. 59

By taking limit in this relation as $k \rightarrow \infty, k \in \mathcal{K}$, and by using the continuity of $H(x, \cdot, \cdot)$ [cf. Assumption 2.4.1(c)], we obtain

$$H(x, \bar{\mu}(x), J^*) \leq H(x, u, J^*), \quad x = 1, \dots, n, \quad u \in U(x).$$

Figure 2.20: P. 59.

The condition of $H(x, \cdot, \cdot)$ being continuous and its relation to the double sequence is elaborated in the note on Assumption 2.4.1 (c), P. 58, [Abstract DP] 2nd Edition.

2.4.1 Approximate Policy Iteration

P. 60

Proposition 2.4.4: ☐ Let the monotonicity and contraction Assumptions 2.1.1 and 2.1.2 hold. Let J , $\bar{\mu}$, and μ satisfy

$$\|J - J_\mu\| \leq \delta, \quad \|T_{\bar{\mu}}J - TJ\| \leq \epsilon,$$

where δ and ϵ are some scalars. Then

$$\|J_{\bar{\mu}} - J^*\| \leq \alpha \|J_\mu - J^*\| + \frac{\epsilon + 2\alpha\delta}{1 - \alpha}. \quad (2.25)$$

Figure 2.21: P. 60.

In the Proof of Proposition 2.4.4 shown below, monotonicity Assumption 2.1.1 is not needed, then why is it stated in the proposition?

Prof. Bertsekas: *the monotonicity assumption is not needed for the proposition as stated. However, monotonicity is an essential assumption for PI to have the fundamental policy improvement property, so I think it's better not to confuse the reader by removing it from the statement of the proposition.*

P. 61

☐ Combining this relation with Eq. (2.28), yields

$$J_{\bar{\mu}} - J^* \leq \alpha \|J_\mu - J^*\| v + \frac{\alpha(\epsilon + 2\alpha\delta)}{1 - \alpha} v + (\epsilon + \alpha\delta)\epsilon = \alpha \|J_\mu - J^*\| v + \frac{\epsilon + 2\alpha\delta}{1 - \alpha} v,$$

which is equivalent to the desired relation (2.25). **Q.E.D.**

Figure 2.22: P. 61 (1).

Here we have implicitly utilized Prop. 2.1.2, P. 43, [Abstract DP] 2nd Edition, which states that $J_{\bar{\mu}} - J^* \geq 0$. Therefore, we do not need to check the boundedness of $J^* - J_{\bar{\mu}}$.

P. 61

Proof of Prop. 2.4.3: Applying Prop. 2.4.4, we have

$$\|J_{\mu^{k+1}} - J^*\| \leq \alpha \|J_{\mu^k} - J^*\| + \frac{\epsilon + 2\alpha\delta}{1 - \alpha},$$

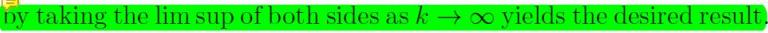
which  by taking the lim sup of both sides as $k \rightarrow \infty$ yields the desired result. **Q.E.D.**

Figure 2.23: P. 61 (2).

Lemma (P. 61). *Given real valued sequence $\{a_n\}$, real-valued constant $\gamma \in (0, 1)$, and real-valued constant β , if*

$$a_{n+1} \leq \gamma a_n + \beta,$$

then $\sup_{k \geq n} a_k < \infty$ and $\limsup_{n \rightarrow \infty} a_n < \infty$.

Proof. $a_1 \leq \gamma a_0 + \beta$, then $a_2 \leq \gamma^2 a_0 + (\gamma + 1)\beta$. By induction, we have

$$a_n \leq \gamma^n a_0 + \frac{1 - \gamma^n}{1 - \gamma} \beta.$$

Since the sequence $\{\gamma^n a_0 + \frac{1 - \gamma^n}{1 - \gamma} \beta\}$ is convergent, then it is also bounded above, therefore $\sup_{k \geq n} a_k < \infty$. Since $\{\sup_{k \geq n} a_k\}_{n=0}^{\infty}$ is nonincreasing, then $\limsup_{n \rightarrow \infty} a_n \leq \sup_{k \geq 0} a_k < \infty$. \square

By [Lemma P. 61], $\{\|J_{\mu^n} - J^*\|\}$ is upper-bounded by positive sequence $\{\alpha^n \|J_{\mu^0} - J^*\| + \frac{1 - \alpha^n}{1 - \alpha} \beta\}$ where $\beta = \frac{\epsilon + 2\alpha\delta}{1 - \alpha}$, therefore, $\limsup \|J_{\mu^n} - J^*\| \in [0, \infty)$, then taking limit superior on both sides of the highlighted part, the limits involved are all real numbers.

2.4.2 Approximate Policy Iteration Where Policies Converge

None.

2.5 Optimistic Policy Iteration and λ -Policy Iteration

P. 64

where J_0 is an initial function in $\mathcal{B}(X)$, and for any policy μ and scalar $\lambda \in (0, 1)$, $T_\mu^{(\lambda)}$ is the multistep mapping defined by

$$T_\mu^{(\lambda)} J = (1 - \lambda) J + \lambda \sum_{\ell=0}^{\infty} \lambda^\ell T_\mu^{\ell+1} J, \quad J \in \mathcal{B}(X)$$

Figure 2.24: P. 64.

The well-definedness of $T_\mu^{(\lambda)}$ is ensured by the following theorems. Their proofs is in [Hand Note 3].

Lemma (P. 64). *Given a real valued sequence $\{a_\ell\}$ and assume that the sequence $\{\sum_{\ell=1}^n a_\ell\}_{n=1}^\infty$ converges with $\sum_{\ell=1}^\infty a_\ell \in \mathbb{R}$. Prove that*

$$\left| \sum_{\ell=1}^\infty a_\ell \right| \leq \sum_{\ell=1}^\infty |a_\ell|. \quad (2.8)$$

Theorem (1, P. 64). *Let the set of mappings $T_\mu : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$, $\mu \in \mathcal{M}$ satisfy Assumption 2.1.2. Consider the mappings $T_\mu^{(w)} : \mathcal{B}(X) \rightarrow \mathcal{R}(X)$ defined by*

$$(T_\mu^{(w)} J)(x) = \sum_{\ell=1}^\infty w_\ell(x) (T_\mu^\ell J)(x), \quad x \in X, J \in \mathcal{B}(X), \quad (2.9)$$

where $w_\ell(x)$ are nonnegative scalars such that for all $x \in X$,

$$\sum_{\ell=1}^\infty w_\ell(x) = 1.$$

Prove that the mapping $T_\mu^{(w)}$ is well defined; namely for all $x \in X$, $J \in \mathcal{B}(X)$, the sequence

$$\left\{ \sum_{\ell=1}^n w_\ell(x) (T_\mu^\ell J)(x) \right\}_{n=1}^\infty \quad (2.10)$$

converges with a limit in \mathbb{R} .

Theorem (2, P. 64). *Let the set of mappings $T_\mu : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$, $\mu \in \mathcal{M}$ satisfy Assumption 2.1.2. Consider the mappings $T_\mu^{(w)} : \mathcal{B}(X) \rightarrow \mathcal{R}(X)$ defined in Eq. (2.9). Prove that $T_\mu^{(w)} \mathcal{B}(X) \subset \mathcal{B}(X)$, namely, $T_\mu^{(w)} : \mathcal{B}(X) \rightarrow \mathcal{R}(X)$ is in fact $T_\mu^{(w)} : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$; and $T_\mu^{(w)}$ is a contraction.*

2.5. OPTIMISTIC POLICY ITERATION AND λ -POLICY ITERATION 29

The following result shows that the operator $T_\mu^{(\lambda)}$ defined point-wise is no difference compared one defined by convergence in norm.

Theorem (3, P. 64). *Consider sequence $\{T_\mu^{(\lambda_n)} J\}$ defined by*

$$T_\mu^{(\lambda_n)} J = (1 - \lambda) \sum_{\ell=1}^n \lambda^{\ell-1} T_\mu^\ell J.$$

The sequence $\{T_\mu^{(\lambda_n)} J\}$ converges to some element $T_\mu^{(\lambda_\infty)} J \in \mathcal{B}(X)$. In addition, it coincides with the limit defined point-wise, viz., $T_\mu^{(\lambda_\infty)} J = T_\mu^{(\lambda)} J$.

Note that the above result does not stand for the more general operator $T_\mu^{(w)}$, when X has infinite cardinality. The following is an example.

Example 2.5.1. Given $X = \{1, 2, \dots\}$ and define $w_\ell(x)$ as

$$w_\ell(x) = 0, \ell \leq x, \quad \sum_{\ell=x+1}^{\infty} w_\ell(x) = 1.$$

Further we assume that $v(x) = x$. Define $T_\mu : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$ as

$$(T_\mu J)(x) = (1 - \alpha)x + \alpha J(x).$$

Then one can verify that $J_\mu(x) = x$. Then consider sequence $\{T_\mu^{(w_n)} J_\mu\}$ defined point-wise as

$$(T_\mu^{(w_n)} J_\mu)(x) = \sum_{\ell=1}^n w_\ell(x) (T_\mu^\ell J_\mu)(x).$$

which can be verified to belong to $\mathcal{B}(X)$. Then $\forall n$, it holds that

$$\begin{aligned} \|T_\mu^{(w_n)} J_\mu - J_\mu\| &= \sup_{x \in X} \frac{|\sum_{\ell=1}^n w_\ell(x) (T_\mu^\ell J_\mu)(x) - J_\mu(x)|}{v(x)} \\ &= \sup_{x \in X} \frac{|\sum_{\ell=n+1}^{\infty} w_\ell(x) J_\mu(x)|}{v(x)}. \end{aligned}$$

Since $\forall n, \exists x$ such that $x > n$. Therefore, we have $\|T_\mu^{(w_n)} J_\mu - J_\mu\| = 1$ for all n . This implies the sequence does not converge in norm. Otherwise, its limit in norm at all x would have same values as $J_\mu(x)$.

2.5.1 Convergence of Optimistic Policy Iteration

P. 66

Lemma 2.5.2: Let the monotonicity and contraction Assumptions 2.1.1 and 2.1.2 hold, let $J \in \mathcal{B}(X)$ and $c \geq 0$ satisfy

$$J \geq TJ - cv,$$

and let $\mu \in \mathcal{M}$ be such that $T_\mu J = TJ$. Then for all $k > 0$, we have

$$TJ \geq T_\mu^k J - \frac{\alpha}{1-\alpha} cv. \quad (2.38)$$

and

$$T_\mu^k J \geq T(T_\mu^k J) - \alpha^k cv. \quad (2.39)$$

Figure 2.25: P. 66 (1).

Due to Eq. (2.11), this lower bound here is never tight.

P. 66

$$TJ = T_\mu J \geq T_\mu^k J - \sum_{j=1}^{k-1} \alpha^j cv = T_\mu^k J - \frac{\alpha - \alpha^k}{1-\alpha} cv \geq T_\mu^k J - \frac{\alpha}{1-\alpha} cv.$$

Figure 2.26: P. 66 (2).

The highlighted part Eq. (2.11) is never tight for any finite k and $\alpha \in (0, 1)$.

$$T_\mu^k J - \frac{\alpha - \alpha^k}{1-\alpha} cv \geq T_\mu^k J - \frac{\alpha}{1-\alpha} cv \quad (2.11)$$

P. 67

Then for all $k \geq 0$,

$$TJ_k + \frac{\alpha}{1-\alpha} \beta_k cv \geq J_{k+1} \geq TJ_{k+1} - \beta_{k+1} cv, \quad (2.41)$$

Figure 2.27: P. 67.

This part of the bound is not tight since in the proof, Eq. (2.38) is used.

P. 68

Then for all $k \geq 0$,

$$J_k + \frac{\alpha^k}{1-\alpha}cv \geq J_k + \frac{\beta_k}{1-\alpha}cv \geq J^* \geq J_k - \frac{(k+1)\alpha^k}{1-\alpha}cv, \quad (2.44)$$

where β_k is defined by Eq. (2.42).

Figure 2.28: P. 68.

Since Eq. (2.38) is used, so this bound is not tight.

P. 69

Proof of Props. 2.5.1 and 2.5.2: Let c be a scalar satisfying Eq. (2.43). Then the error bounds (2.44) show that $\lim_{k \rightarrow \infty} \|J_k - J^*\| = 0$, i.e., the first part of Prop. 2.5.1. To show the second part (finite termination when the

Figure 2.29: P. 69 (1).

By Eq. (2.44), $J^* - J_k \leq \frac{\alpha^k}{1-\alpha}cv$, $J_k - J^* \leq \frac{(k+1)\alpha^k}{1-\alpha}cv$, we have

$$|J_k - J^*| \leq \frac{(k+1)\alpha^k}{1-\alpha}cv \implies \|J_k - J^*\| \leq \frac{(k+1)\alpha^k}{1-\alpha}.$$

P. 69 (20190814)

Proof of Props. 2.5.1 and 2.5.2: Let c be a scalar satisfying Eq. (2.43). Then the error bounds (2.44) show that $\lim_{k \rightarrow \infty} \|J_k - J^*\| = 0$, i.e., the first part of Prop. 2.5.1. To show the second part (finite termination when the number of policies is finite), let $\widehat{\mathcal{M}}$ be the finite set of nonoptimal policies. Then there exists $\epsilon > 0$ such that $\|T_{\widehat{\mu}}J^* - TJ^*\| > \epsilon$ for all $\widehat{\mu} \in \widehat{\mathcal{M}}$, which implies that $\|T_{\widehat{\mu}}J_k - TJ_k\| > \epsilon$ for all $\widehat{\mu} \in \widehat{\mathcal{M}}$ and k sufficiently large. This implies that $\mu^k \notin \widehat{\mathcal{M}}$ for all k sufficiently large. The proof of Prop. 2.5.2 follows using the compactness and continuity Assumption 2.4.1, and the convergence argument of Prop. 2.4.2. **Q.E.D.**

Figure 2.30: P. 69 (20190814).

Since $\widehat{\mathcal{M}}$ is finite, denote $\delta = \min_{\hat{\mu} \in \widehat{\mathcal{M}}} \|T_{\hat{\mu}} J^* - T J^*\|$, then $\forall \varepsilon \in (0, \delta)$ and $\forall \hat{\mu} \in \widehat{\mathcal{M}}$, it holds that $\|T_{\hat{\mu}} J^* - T J^*\| > \varepsilon$. Then we have

$$\begin{aligned} \|T_{\hat{\mu}} J_k - T J_k\| &= \|T_{\hat{\mu}} J_k - T_{\hat{\mu}} J^* + T_{\hat{\mu}} J^* - T J^* + T J^* - T J_k\| \\ &\geq \|T_{\hat{\mu}} J^* - T J^*\| - \|T_{\hat{\mu}} J_k - T_{\hat{\mu}} J^*\| - \|T J^* - T J_k\| \\ &\geq \|T_{\hat{\mu}} J^* - T J^*\| - 2\alpha \|J^* - J_k\| \\ &\geq \delta - 2\alpha \|J^* - J_k\|. \end{aligned}$$

Since $J_k \rightarrow J^*$ in norm, then $\exists N$ such that $\forall k > N$, $\delta - \alpha \|J^* - J_k\| > \varepsilon$.

P. 69

In comparing the bounds (2.47) and (2.48), we should also take into account the associated overhead for a single iteration of each method: optimistic PI requires at iteration k a single application of T and $m_k - 1$ applications of T_{μ^k} (each being less time-consuming than an application of T), while VI requires a single application of T . it can then be seen that the upper bound for optimistic PI is better than the one for VI (same bound for less overhead), while the lower bound for optimistic PI is worse than the one for VI (worse bound for more overhead). This suggests that the choice of the initial condition J_0 is important in optimistic PI, and in particular it is preferable to have $J_0 \geq T J_0$ (implying convergence to J^* from above) rather than $J_0 \leq T J_0$ (implying convergence to J^* from below). This is consistent with the results of other works, which indicate that the convergence properties of the method are fragile when the condition $J_0 \geq T J_0$ does not hold (see [WiB93], [BeT96], [BeY10], [BeY12], [YuB13a]).

Figure 2.31: P. 69 (2).

Regarding bounds given Eq. (2.47) here, the 'upper bound for optimistic PI' is refers to

$$J^* \leq J_k + \frac{\alpha^{m_0 + \dots + m_k}}{1 - \alpha} cv \iff J^* - \frac{\alpha^{m_0 + \dots + m_k}}{1 - \alpha} cv \leq J_k \quad (2.12)$$

and the 'lower bound for optimistic PI' is refers to

$$J_k - \frac{(k+1)\alpha^k}{1 - \alpha} cv \leq J^* \iff J_k \leq J^* + \frac{(k+1)\alpha^k}{1 - \alpha} cv. \quad (2.13)$$

Given $J_0 \geq T J_0$, we have

$$J_0 \geq T J_0 = T_{\mu^0} J_0 \geq T_{\mu^0}^{m_0 - 1} J_0 \geq T_{\mu^0}^{m_0} J_0 = J_1 \geq T T_{\mu^0}^{m_0 - 1} J_0 \geq T T_{\mu^0}^{m_0} J_0 = T J_1,$$

namely, given $J_0 \geq T J_0$, it holds for all k that $J_k \geq T J_k$ and $J_k \geq J^*$. Therefore, the bound that actually regulate the error is (2.13), which is still worse than VI, and the bound (2.12) is automatically fulfilled. On the other

hand, if $J_0 \leq TJ_0$, it is not true that for all k that $J_k \leq TJ_k$ and $J_k \leq J^*$, one special example is given in Fig. 2.5.1, P. 64, [Abstract DP] 2nd Edition, where $J_0 \leq TJ_0$ but $J_1 \geq TJ_1$, which means eventually it converges from above.

2.5.2 Approximate Optimistic Policy Iteration

P. 70

$$M(y) = \sup_{x \in X} \frac{y(x)}{v(x)}.$$

Figure 2.32: P. 70.

Lemma (P. 70). *Given $f, g, h \in \mathcal{B}(X)$, show that*

$$\sup_{x \in X} (f - g) \leq \sup_{x \in X} (f - h) + \sup_{x \in X} (h - g). \quad (2.14)$$

In particular, if g is constant 0, it holds

$$\sup_{x \in X} f \leq \sup_{x \in X} (f - h) + \sup_{x \in X} h. \quad (2.15)$$

Proof. $\forall x \in X$, it holds that

$$(f - h)(x) \leq \sup_{x \in X} (f - h), \quad (h - g)(x) \leq \sup_{x \in X} (h - g),$$

therefore we have

$$(f - g)(x) = (f - h + h - g)(x) \leq \sup_{x \in X} (f - h) + \sup_{x \in X} (h - g),$$

take supremum on both sides and we get the desired result. \square

Theorem (P. 70). *M is continuous with respect to the weighted supremum norm $\|\cdot\|$; namely, given $f_n, f \in \mathcal{B}(X)$, if $\lim_{n \rightarrow \infty} \|f_n - f\| = 0$, then $\lim_{n \rightarrow \infty} M(f_n) = M(f)$.*

Proof. By [Lemma P. 70], $M(f_n) \leq M(f_n - f) + M(f)$, we have

$$\limsup_{n \rightarrow \infty} M(f_n) \leq \limsup_{n \rightarrow \infty} M(f_n - f) + M(f) \leq \limsup_{n \rightarrow \infty} \|f_n - f\| + M(f) = M(f),$$

where we have applied [Lemmas P. 42]. On the other hand, due to definition of $M(\cdot)$,

$$M(f_n) = \sup_{x \in X} \frac{f_n(x)}{v(x)} = \sup_{x \in X} \frac{f(x) + f_n(x) - f(x)}{v(x)}. \quad (2.16)$$

In addition, we have

$$\begin{aligned}
\frac{f(x) + f_n(x) - f(x)}{v(x)} &\geq \frac{f(x)}{v(x)} - \frac{|f_n(x) - f(x)|}{v(x)} \\
&\geq \frac{f(x)}{v(x)} - \sup_{y \in X} \frac{|f_n(y) - f(y)|}{v(y)} \\
&= \frac{f(x)}{v(x)} - \|f_n - f\|.
\end{aligned} \tag{2.17}$$

Taking Eq. (2.17) into Eq. (2.16), we have

$$M(f_n) \geq \sup_{x \in X} \frac{f(x)}{v(x)} - \|f_n - f\| = M(f) - \|f_n - f\|.$$

Taking limit inferior on both sides, we have

$$\liminf_{n \rightarrow \infty} M(f_n) \geq M(f) - \limsup_{n \rightarrow \infty} \|f_n - f\| = M(f),$$

which conclude the proof. \square

P. 71

Lemma (P. 71). *Given real sequences $\{r_n\}$, $\{s_n\}$, $\{t_n\}$ and $\{a_n\}$, assume the following conditions hold for all $n \geq 1$:*

$$r_n \leq \alpha_r r_{n-1} + \beta; \tag{2.18}$$

$$s_n \leq \alpha_s r_n; \tag{2.19}$$

$$t_n \leq \alpha_t t_{n-1} + \gamma r_n + \delta; \tag{2.20}$$

$$a_n = s_n + t_n, \quad a_n \geq \zeta; \tag{2.21}$$

where the constants $\alpha_r, \alpha_t \in (0, 1)$, the constants α_s, γ are positive real values, and the constants β, δ, ζ are real values. Then the limit superiors $\limsup_{n \rightarrow \infty} r_n$, $\limsup_{n \rightarrow \infty} s_n$ and $\limsup_{n \rightarrow \infty} t_n$ are all real values.

Proof. By [Lemma P. 61], $\forall n \sup_{k \geq n} r_k < \infty$ and the sequence $\{r_n\}$ is bounded by the sequence $\{\alpha^n r_0 + \frac{1-\alpha^n}{1-\alpha} \beta\}$. Since $\{\alpha^n r_0 + \frac{1-\alpha^n}{1-\alpha} \beta\}$ is convergent, then it is bounded by some constant, denoted as M_r . Therefore, $\sup_{k \geq n} r_k \leq M_r$. As a result, we have $\sup_{k \geq n} s_k \leq M_s$ where $M_s = \alpha_s M_r$. Apply the upper bound on r_n to Eq. (2.20), we have

$$t_n \leq \alpha_t t_{n-1} + \gamma M_r + \delta,$$

and by [Lemma P. 61], we have $\{t_n\}$ is bounded by some constant M_t . In what follows, we show that Eq. (2.21) guarantees that the limit superiors are reals. Adding Eq. (2.20) and (2.19), and by Eq. (2.21), we have

$$\zeta \leq a_n = s_n + t_n \leq (\alpha_s + \gamma) r_n + \alpha_t M_t + \delta,$$

which indicates $\{r_n\}$ is lower bounded since $\alpha_s + \gamma$ is some positive constant. Similarly, by Eq. (2.21),

$$t_n = a_n - s_n \geq \zeta - M_s, \quad s_n = a_n - t_n \geq \zeta - M_t.$$

Therefore, their limit superiors are real. \square

Alternatively, we can prove by contradiction. Once we get M_r, M_s and M_t , note that the sequence $\{\sup_{k \geq n} r_k\}$ is monotonically nonincreasing, then if it is unbounded below, given $-(M_t + \varepsilon)/\alpha_s, \exists N$ such that $\sup_{k \geq N} r_k \leq (\zeta - M_t - \varepsilon)/\alpha_s$. By Eq. (2.19), we have $\forall n \geq N, s_k \leq \zeta - M_t - \varepsilon$. Therefore, we have $t_k + s_k \leq \zeta - \varepsilon$, which contradicts Eq. (2.21). Therefore, $\{\sup_{k \geq n} r_k\}$ is bounded below. Repeat the same arguments, we can have that $\{\sup_{k \geq n} s_k\}$ and $\{\sup_{k \geq n} t_k\}$ are bounded below. Therefore, their limit superiors are real.

P. 74

Given Assumption 2.1.2 holds, namely $\|T_\mu J_k - T_\mu J\| \leq \alpha \|J_k - J\|$, and $\|T J_k - T J\| \leq \alpha \|J_k - J\|$ where $\alpha \in (0, 1)$. Then for all sequence $\{J_k\} \subset \mathcal{B}(X)$ such that $\|J_k - J\| \rightarrow 0$ where $J \in \mathcal{B}(X)$, we have $\forall k \|T_\mu J_k - T_\mu J\| \leq \alpha \|J_k - J\|$, and $\|T J_k - T J\| \leq \alpha \|J_k - J\|$. Namely, with Assumption 2.1.2 true, we have

$$\begin{aligned} \|J_k - J\| \rightarrow 0 &\implies \|T_\mu J_k - T_\mu J\| \rightarrow 0; \\ \|J_k - J\| \rightarrow 0 &\implies \|T J_k - T J\| \rightarrow 0. \end{aligned}$$

P. 75

Denote the second part of Prop. 2.5.4 as property P , where

$$P = (\exists \varepsilon > 0)(\forall \|J - J^*\| < \varepsilon)(T_\mu J = T J \implies \mu \in \mathcal{M}^*),$$

then we have

$$\begin{aligned} \neg P &= (\forall \varepsilon > 0)(\exists \|J - J^*\| < \varepsilon)(T_\mu J = T J \not\Rightarrow \mu \in \mathcal{M}^*) \\ &= (\forall \varepsilon > 0)(\exists \|J - J^*\| < \varepsilon, \exists \mu \in \mathcal{M} \setminus \mathcal{M}^*)(T_\mu J = T J). \end{aligned}$$

Since \mathcal{M} is finite, so as $\mathcal{M} \setminus \mathcal{M}^*$, then given the sequence $\{\varepsilon_k\}$ and its corresponding $\{J_k\}$ and $\{\mu_k\}$, there must be some $\bar{\mu} \in \{\mu_k\}$ which repeated infinite times.

P. 77 (20190902)

As shown in [Theorem 3 P. 64], when the complete space $\mathcal{F}(X) = \mathcal{B}(X)$ and the norm is weighted sup-norm, there is no difference between interpreting $T_\mu^{(\lambda)} J$ as a function of x defined as point-wise limit

$$(T_\mu^{(\lambda)} J)(x) = (1 - \lambda) \sum_{\ell=1}^{\infty} \lambda^{\ell-1} (T_\mu^\ell J)(x), \quad x \in X, J \in \mathcal{B}(X), \quad (2.22)$$

or as an element of $\mathcal{F}(X)$ which is the limit of the convergent sequence $\{T_\mu^{(\lambda_n)} J\}$ in $\mathcal{F}(X)$ given as

$$\lim_{n \rightarrow \infty} \|T_\mu^{(\lambda_n)} J - T_\mu^{(\lambda)} J\| = 0 \quad (2.23)$$

where

$$T_\mu^{(\lambda_n)} J = (1 - \lambda) \sum_{\ell=1}^n \lambda^{\ell-1} T_\mu^\ell J.$$

However, since the nature of $\mathcal{F}(X)$ and the norm $\|\cdot\|$ are left unspecified, $T_\mu^{(\lambda)} J$ shall be interpreted as the in Eq. (2.23). In what follows, we repeat the first part of [Theorem 3 P. 64], which shows that $\{T_\mu^{(\lambda_n)} J\}$ is convergent in $\mathcal{F}(X)$ and therefore $T_\mu^{(\lambda)}$ as its limit is well-defined.

Theorem (P. 77). *Let Assumption 2.5.2 (b) hold. Then $\forall J \in \mathcal{F}(X)$, the sequence $\{T_\mu^{(\lambda_n)} J\}$ defined by*

$$T_\mu^{(\lambda_n)} J = (1 - \lambda) \sum_{\ell=1}^n \lambda^{\ell-1} T_\mu^\ell J$$

is convergent.

Proof. Since $\lim_{n \rightarrow \infty} \|T_\mu^n J - J_\mu\| = 0$, we have $\lim_{n \rightarrow \infty} \|T_\mu^n J\| = \|J_\mu\|$ [cf. Theorem P. 42]. Therefore $\{\|T_\mu^n J\|\}$ is bounded. Denote its bound as M_μ .

Therefore, $\forall \varepsilon, \exists N$ such that $\forall k$

$$\begin{aligned}
& \|T_\mu^{(\lambda_N)} J - T_\mu^{(\lambda_{N+k})} J\| \\
&= \|(1-\lambda) \sum_{\ell=1}^N \lambda^{\ell-1} T_\mu^\ell J - (1-\lambda) \sum_{\ell=1}^{N+k} \lambda^{\ell-1} T_\mu^\ell J\| \\
&= \|(1-\lambda) \sum_{\ell=N+1}^{N+k} \lambda^{\ell-1} T_\mu^\ell J\| \\
&\leq (1-\lambda) \sum_{\ell=N+1}^{N+k} \lambda^{\ell-1} \|T_\mu^\ell J\| \\
&\leq (1-\lambda) \sum_{\ell=N+1}^{N+k} \lambda^{\ell-1} M_\mu \\
&\leq \lambda^N M_\mu \\
&\leq \varepsilon,
\end{aligned}$$

which implies $\{T_\mu^{(\lambda_n)} J\}$ is Cauchy. Since $\mathcal{F}(X)$ is complete, then it is also convergent. \square

P. 87 (20190902)

The following lemma shows the relation between the Cartesian product of $\mathcal{B}(X_\ell)$ and $\mathcal{B}(X)$.

Lemma (1, P. 87). *Given processor index set I being finite, and $\{X_\ell\}_{\ell \in I}$ is a partition of X , then it holds that*

$$\prod_{\ell \in I} \mathcal{B}(X_\ell) = \mathcal{B}(X). \quad (2.24)$$

Proof. First we show that $\mathcal{B}(X) \subseteq \prod_{\ell \in I} \mathcal{B}(X_\ell)$. Given $J \in \mathcal{B}(X)$ and denoted as J_ℓ the restriction of J on X_ℓ . Then we have

$$\sup_{x \in X_\ell} \frac{|J_\ell(x)|}{v(x)} \leq \sup_{x \in X} \frac{|J(x)|}{v(x)} = \|J\| < \infty, \forall \ell$$

where the first inequality is due to that the supremum of an upper bounded real set is no less than the supremum of its subset. Therefore, we have $J_\ell \in \mathcal{B}(X_\ell) \forall \ell$ and consequently $J \in \prod_{\ell \in I} \mathcal{B}(X_\ell)$.

On the other hand, given $J \in \prod_{\ell \in I} \mathcal{B}(X_\ell)$, and denote as M_ℓ the bound of $J_\ell \in \mathcal{B}(X_\ell)$, then we have

$$\frac{|J(x)|}{v(x)} \leq \sum_{\ell \in I} M_\ell \chi_{X_\ell}(x)$$

where $\chi_{X_\ell}(\cdot)$ is the indicator functions defined on X . Then take supremum on both sides of the equation, we have

$$\sup_{x \in X} \frac{|J(x)|}{v(x)} \leq \sup_{x \in X} \sum_{\ell \in I} M_\ell \chi_{X_\ell}(x) = \sup_{\ell \in I} \{M_\ell\} = \max_{\ell \in I} \{M_\ell\} < \infty.$$

Note that I being finite is needed. Otherwise, the bound of $\prod_{\ell \in I} J_\ell$ is $\sup_{\ell \in I} \{M_\ell\}$, which may be ∞ . \square

The following lemma proves the relation between the Cartesian product and set intersection under different construction.

Lemma (2, P. 87). *Given policy set \mathcal{M} and processor index set I both being finite, and $\{X_\ell\}_{\ell \in I}$ is a partition of X , then it holds that*

$$\prod_{\ell \in I} \mathcal{J}_{|X_\ell}(\alpha) = \bigcap_{\ell \in I} \mathcal{J}_\ell(\alpha) \quad (2.25)$$

where

$$\mathcal{J}_{|X_\ell}(\alpha) = \left\{ J_\ell \in \mathcal{B}(X_\ell) \left| \max_{\mu \in \mathcal{M}} \sup_{x \in X_\ell} \frac{|J_\ell(x) - J_\mu(x)|}{v(x)} \leq \alpha \right. \right\}, \quad (2.26)$$

$$\mathcal{J}_\ell(\alpha) = \left\{ J \in \mathcal{B}(X) \left| \max_{\mu \in \mathcal{M}} \sup_{x \in X_\ell} \frac{|J(x) - J_\mu(x)|}{v(x)} \leq \alpha \right. \right\}. \quad (2.27)$$

Proof. Note that we need to apply the result of [Lemma 1, P. 87] that

$$\prod_{\ell \in I} \mathcal{B}(X_\ell) = \mathcal{B}(X)$$

to establish that the underline sets $\prod_{\ell \in I} \mathcal{B}(X_\ell)$ and $\mathcal{B}(X)$ are the same. With that fact in mind, we first show that $\prod_{\ell \in I} \mathcal{J}_{|X_\ell}(\alpha) \subseteq \bigcap_{\ell \in I} \mathcal{J}_\ell(\alpha)$. Indeed, for $J \in \prod_{\ell \in I} \mathcal{J}_{|X_\ell}(\alpha)$, it holds $\forall \ell \in I, \forall x \in X_\ell, \forall \mu \in \mathcal{M}$, that

$$\frac{|J(x) - J_\mu(x)|}{v(x)} = \frac{|J_\ell(x) - J_\mu(x)|}{v(x)} \leq \max_{\mu \in \mathcal{M}} \sup_{x \in X_\ell} \frac{|J_\ell(x) - J_\mu(x)|}{v(x)} \leq \alpha. \quad (2.28)$$

Take supremum over $x \in X_\ell$ on both sides and then take maximum over $\mu \in \mathcal{M}$ on both sides, and we have $J \in \bigcap_{\ell \in I} \mathcal{J}_\ell(\alpha)$.

On the other hand, given $J \in \bigcap_{\ell \in I} \mathcal{J}_\ell(\alpha)$, it holds $\forall \ell \in I, \forall x \in X_\ell, \forall \mu \in \mathcal{M}$, that

$$\frac{|J_\ell(x) - J_\mu(x)|}{v(x)} = \frac{|J(x) - J_\mu(x)|}{v(x)} \leq \max_{\mu \in \mathcal{M}} \sup_{x \in X_\ell} \frac{|J(x) - J_\mu(x)|}{v(x)} \leq \alpha, \quad (2.29)$$

which implies $J_\ell \in \mathcal{J}_{|X_\ell}(\alpha) \forall \ell \in I$, and this concludes the proof. \square

Lemma (3, P. 87). *Given policy set \mathcal{M} and processor index set I both being finite, and $\{X_\ell\}_{\ell \in I}$ is a partition of X , and given $\mathcal{J}(\alpha)$ and $\mathcal{J}_\ell(\alpha)$ as*

$$\begin{aligned}\mathcal{J}(\alpha) &= \left\{ J \in \mathcal{B}(X) \mid \max_{\mu \in \mathcal{M}} \|J - J_\mu\| \leq \alpha \right\}, \\ \mathcal{J}_\ell(\alpha) &= \left\{ J \in \mathcal{B}(X) \mid \max_{\mu \in \mathcal{M}} \sup_{x \in X_\ell} \frac{|J(x) - J_\mu(x)|}{v(x)} \leq \alpha \right\}.\end{aligned}$$

Then it holds that

$$\mathcal{J}(\alpha) = \bigcap_{\ell \in I} \mathcal{J}_\ell(\alpha). \quad (2.30)$$

Proof. Here we introduce two kinds of sets defined as

$$\begin{aligned}\mathcal{J}^\mu(\alpha) &= \left\{ J \in \mathcal{B}(X) \mid \|J - J_\mu\| \leq \alpha \right\}, \\ \mathcal{J}_\ell^\mu(\alpha) &= \left\{ J \in \mathcal{B}(X) \mid \sup_{x \in X_\ell} \frac{|J(x) - J_\mu(x)|}{v(x)} \leq \alpha \right\}.\end{aligned}$$

Then one can verify that between those two sets, it holds that

$$\mathcal{J}^\mu(\alpha) = \bigcap_{\ell \in I} \mathcal{J}_\ell^\mu(\alpha), \quad \forall \mu \in \mathcal{M}.$$

In addition, one may verify that

$$\mathcal{J}(\alpha) = \bigcap_{\mu \in \mathcal{M}} \mathcal{J}^\mu(\alpha), \quad \mathcal{J}_\ell(\alpha) = \bigcap_{\mu \in \mathcal{M}} \mathcal{J}_\ell^\mu(\alpha). \quad (2.31)$$

Therefore, we have

$$\begin{aligned}\mathcal{J}(\alpha) &= \bigcap_{\mu \in \mathcal{M}} \mathcal{J}^\mu(\alpha) \\ &= \bigcap_{\mu \in \mathcal{M}} \bigcap_{\ell \in I} \mathcal{J}_\ell^\mu(\alpha) \\ &= \bigcap_{\ell \in I} \bigcap_{\mu \in \mathcal{M}} \mathcal{J}_\ell^\mu(\alpha) \\ &= \bigcap_{\ell \in I} \mathcal{J}_\ell(\alpha)\end{aligned}$$

□

Theorem (P. 87). *Given policy set \mathcal{M} and processor index set I both being finite, and $\{X_\ell\}_{\ell \in I}$ is a partition of X , and given $\mathcal{J}(\alpha)$ and $\mathcal{J}_{|X_\ell}(\alpha)$ as*

$$\begin{aligned}\mathcal{J}(\alpha) &= \left\{ J \in \mathcal{B}(X) \mid \max_{\mu \in \mathcal{M}} \|J - J_\mu\| \leq \alpha \right\}, \\ \mathcal{J}_{|X_\ell}(\alpha) &= \left\{ J_\ell \in \mathcal{B}(X_\ell) \mid \max_{\mu \in \mathcal{M}} \sup_{x \in X_\ell} \frac{|J_\ell(x) - J_\mu(x)|}{v(x)} \leq \alpha \right\}.\end{aligned}$$

Then it holds that

$$\mathcal{J}(\alpha) = \prod_{\ell \in I} \mathcal{J}_{|X_\ell}(\alpha). \quad (2.32)$$

Proof. Combining [Lemmas 1, 2, 3, P. 87], we get the desired result. \square

Remark. As noted in the proof of [Lemma 1, P. 87], it is needed to have I being finite in order to have the underline sets $\prod_{\ell \in I} \mathcal{B}(X_\ell)$ and $\mathcal{B}(X)$ being the same. However, as far as the proofs being concerned, it is not needed to have \mathcal{M} being finite. The steps related in the proofs are (2.28), (2.29), and (2.31), which all holds if all $\max_{\mu \in \mathcal{M}}$ involved are replaced by $\sup_{\mu \in \mathcal{M}}$ properly.

P. 87 (20190902)

With the Box condition proved, we can show that the sequence $\{J^t\}$ generated by the algorithm is bounded. First, we denote Δ_1 and Δ_2 as

$$\begin{aligned} \Delta_1 &= \max_{\mu \in \mathcal{M}} \|J^0 - J_\mu\|, \\ \Delta_2 &= \max_{\mu, \mu' \in \mathcal{M}} \|J_\mu - J_{\mu'}\|, \end{aligned}$$

and we have $\Delta_1 \leq \max\{\Delta_1, \Delta_2/(1-\alpha)\}$. Denote as \hat{J}^1 the function had the first update is done synchronously, namely all components are updated, rather than only $\forall x \in X_\ell$ where $0 \in \mathcal{R}_\ell \cup \overline{\mathcal{R}}_\ell$, then we have

$$\begin{aligned} \|\hat{J}^1 - J_\mu\| &\leq \|\hat{J}^1 - J_{\mu^0}\| + \|J_{\mu^0} - J_\mu\| \\ &= \|T_{\mu^0} J^0 - J_{\mu^0}\| + \|J_{\mu^0} - J_\mu\| \\ &\leq \alpha \|J^0 - J_{\mu^0}\| + \|J_{\mu^0} - J_\mu\| \\ &\leq \alpha \max_{\mu \in \mathcal{M}} \|J^0 - J_\mu\| + \max_{\mu, \mu' \in \mathcal{M}} \|J_\mu - J_{\mu'}\| \\ &= \alpha \Delta_1 + \Delta_2 \\ &\leq \alpha \max\left\{\Delta_1, \frac{\Delta_2}{1-\alpha}\right\} + \Delta_2 \\ &= \max\left\{\alpha \Delta_1 + \Delta_2, \frac{\Delta_2}{1-\alpha}\right\}. \end{aligned}$$

Take $\max_{\mu \in \mathcal{M}}$ on both sides, we have

$$\max_{\mu \in \mathcal{M}} \|\hat{J}^1 - J_\mu\| \leq \max\left\{\alpha \Delta_1 + \Delta_2, \frac{\Delta_2}{1-\alpha}\right\}.$$

If $\Delta_1 \geq \Delta_2/(1-\alpha)$, we have $\alpha \Delta_1 + \Delta_2 \leq \alpha \Delta_1 + (1-\alpha)\Delta_1 = \Delta_1$, and $\alpha \Delta_1 + \Delta_2 \geq \Delta_2/(1-\alpha)$; otherwise, we have $\alpha \Delta_1 + \Delta_2 < \Delta_2/(1-\alpha)$. With this shown, we have

$$\max_{\mu \in \mathcal{M}} \|\hat{J}^1 - J_\mu\| \leq \max\left\{\Delta_1, \frac{\Delta_2}{1-\alpha}\right\}.$$

Denote $\Delta = \max\{\Delta_1, \Delta_2/(1 - \alpha)\}$. Then we see that $\hat{J}^1, J^0 \in \mathcal{J}(\Delta)$. Now consider the real J^1 , which is given as

$$J^1(x) = \begin{cases} \hat{J}^1(x) & \text{if } x \in X_\ell, 0 \in \mathcal{R}_\ell \cup \overline{\mathcal{R}}_\ell, \\ J^0(x) & \text{o.w..} \end{cases}$$

Then by the box condition, we see that J^1 , which is Cartesian products of portions of J^0 and \hat{J}^1 , is also in the set $\mathcal{J}(\Delta)$. By induction, we can establish that $J^t \in \mathcal{J}(\Delta) \forall t$.

P. 90

Here we show that the norm is well-defined and the corresponding space is complete. To this end, we define a new state space S whose elements are all feasible state control pairs $s = (x, u)$ where $u \in U(x)$. If we denote as S_x the set of state and control pairs where the state is x and the control $u \in U(x)$, viz., the set $\cup_{u \in U(x)}\{(x, u)\}$, then we see that $S_x \cap S_{x'} = \emptyset$ for $x \neq x'$ and

$$S = \bigcup_{x \in X} S_x.$$

Define a positive function $\hat{v} : S \rightarrow \mathbb{R}_+$ where $\hat{v}(s) = v(x)$ with $s = (x, u)$. In addition, given $V(\cdot)$ and $Q(\cdot, \cdot)$, we have corresponding functions $\hat{V} : S \rightarrow \mathbb{R}$ and $\hat{Q} : S \rightarrow \mathbb{R}$ defined as

$$\hat{V}(s) = V(x), \hat{Q}(s) = Q(x, u), \text{ with } s = (x, u). \quad (2.33)$$

Then with the weight \hat{v} and corresponding norm denoted as $\|\cdot\|_{\hat{v}}$, we denote as $\mathcal{B}_1(S)$ the functional space whose elements are functions defined on S to \mathbb{R} with aforementioned weighted sup-norm bounded. Then we have $\mathcal{B}_1(S)$ is complete with respect to $\|\cdot\|_{\hat{v}}$. The proof can be found in P. 329, [Abstract DP], 2nd Edition. Now consider instead functions $W : S \rightarrow \mathbb{R}^2$ of the form

$$W(s) = (\hat{V}(s), \hat{Q}(s)).$$

Consider now the functional space $\mathcal{B}_2(S)$ whose elements are W with the property that $\max\{\|\hat{V}\|_{\hat{v}}, \|\hat{Q}\|_{\hat{v}}\} < \infty$. Now we prove

$$\|W\| = \max\{\|\hat{V}\|_{\hat{v}}, \|\hat{Q}\|_{\hat{v}}\}$$

indeed defines a norm on $\mathcal{B}_2(S)$. We only show the triangular inequality part. The proof for the other two properties are neglected. Given $W_1 = (\hat{V}_1, \hat{Q}_1)$, $W_2 = (\hat{V}_2, \hat{Q}_2)$, we have $W_1 + W_2 = (\hat{V}_1 + \hat{V}_2, \hat{Q}_1 + \hat{Q}_2)$. Then we have

$$\|W_1 + W_2\| = \max\{\|\hat{V}_1 + \hat{V}_2\|_{\hat{v}}, \|\hat{Q}_1 + \hat{Q}_2\|_{\hat{v}}\}.$$

Since we have

$$\begin{aligned} \|\hat{V}_1 + \hat{V}_2\|_{\hat{v}} &\leq \|\hat{V}_1\|_{\hat{v}} + \|\hat{V}_2\|_{\hat{v}} \\ &\leq \max\{\|\hat{V}_1\|_{\hat{v}}, \|\hat{Q}_1\|_{\hat{v}}\} + \max\{\|\hat{V}_2\|_{\hat{v}}, \|\hat{Q}_2\|_{\hat{v}}\} \\ &= \|W_1\| + \|W_2\|, \end{aligned}$$

and similarly

$$\|\hat{Q}_1 + \hat{Q}_2\|_{\hat{v}} \leq \|W_1\| + \|W_2\|.$$

We then have

$$\|W_1 + W_2\| = \max\{\|\hat{V}_1 + \hat{V}_2\|_{\hat{v}}, \|\hat{Q}_1 + \hat{Q}_2\|_{\hat{v}}\} \leq \|W_1\| + \|W_2\|,$$

which confirms in part that the norm $\|\cdot\|$ is well-defined.

Now we proceed to prove $\mathcal{B}_2(S)$ is complete. Given Cauchy sequence $\{W_k\} \subset \mathcal{B}_2(S)$ with respect to $\|\cdot\|$ where $W_k = (\hat{V}_k, \hat{Q}_k)$, we can see that $\{\hat{V}_k\}, \{\hat{Q}_k\} \subset \mathcal{B}_1(S)$ are both Cauchy with respect to $\|\cdot\|_{\hat{v}}$. Note that since $\{W_k\}$ is some arbitrary Cauchy sequence, so \hat{V}_k is more general than the form defined in Eq. (2.33). Since $\mathcal{B}_1(S)$ is complete, $\|\hat{V}_k - \hat{V}^*\|_{\hat{v}} \rightarrow 0$ and $\|\hat{Q}_k - \hat{Q}^*\|_{\hat{v}} \rightarrow 0$ where $\hat{V}^*, \hat{Q}^* \in \mathcal{B}_1(S)$. Denote $W^* = (\hat{V}^*, \hat{Q}^*)$. Then it's easy to see that $W^* \in \mathcal{B}_2(S)$. In addition, since $\forall \varepsilon > 0, \exists K_V(\varepsilon)$ and $\exists K_Q(\varepsilon)$ such that $\|\hat{V}_k - \hat{V}^*\|_{\hat{v}} < \varepsilon \forall k \geq K_V(\varepsilon)$, and $\|\hat{Q}_k - \hat{Q}^*\|_{\hat{v}} < \varepsilon \forall k \geq K_Q(\varepsilon)$, then $\forall k \geq K(\varepsilon) = \max\{K_V(\varepsilon), K_Q(\varepsilon)\}$, $\max\{\|\hat{V}_k - \hat{V}^*\|_{\hat{v}}, \|\hat{Q}_k - \hat{Q}^*\|_{\hat{v}}\} < \varepsilon$, viz.,

$$\|W_k - W^*\| < \varepsilon, \forall k \geq K(\varepsilon),$$

which shows that $\{W_k\}$ is convergent and $\mathcal{B}_2(S)$ therefore is complete.

P. 90

We continue to use the notations introduced above. Note the difference between $\mathcal{B}(X)$ and $\mathcal{B}_1(S)$. If we assume $H(\cdot, \cdot, \cdot)$ is such that $\hat{Q}(s) = Q(x, u) = H(x, u, J)$ is in the space $\mathcal{B}_1(S) \forall J \in \mathcal{B}(X)$, then it implies that $T_\mu J \in \mathcal{B}(X)$ and $TJ \in \mathcal{B}(X) \forall J \in \mathcal{B}(X)$ (part of Assumption 2.1.2). To see this, we denote as S_μ the set $\cup_{x \in X} \{(x, \mu(x))\}$. Then we have

$$\sup_{x \in X} \frac{|(T_\mu J)(x)|}{v(x)} = \sup_{s \in S_\mu} \frac{|\hat{Q}(s)|}{\hat{v}(s)} \leq \sup_{s \in S} \frac{|\hat{Q}(s)|}{\hat{v}(s)} < \infty$$

since $S_\mu \subseteq S$. In addition, we denote as S_x the set $\cup_{u \in U(x)} \{(x, u)\}$, then we have

$$(TJ)(x) = \inf_{u \in U(x)} H(x, u, J) = v(x) \inf_{s \in S_x} \frac{\hat{Q}(s)}{\hat{v}(s)}$$

as $\forall s \in S_x, \hat{v}(s) = v(x)$. Since

$$-\frac{|\hat{Q}(s)|}{\hat{v}(s)} \leq \frac{\hat{Q}(s)}{\hat{v}(s)} \leq \frac{|\hat{Q}(s)|}{\hat{v}(s)} \implies -\sup_{s \in S_x} \frac{|\hat{Q}(s)|}{\hat{v}(s)} \leq \inf_{s \in S_x} \frac{\hat{Q}(s)}{\hat{v}(s)} \leq \sup_{s \in S_x} \frac{|\hat{Q}(s)|}{\hat{v}(s)},$$

we have

$$|(TJ)(x)| = \left| v(x) \inf_{s \in S_x} \frac{\hat{Q}(s)}{\hat{v}(s)} \right| \leq v(x) \sup_{s \in S_x} \frac{|\hat{Q}(s)|}{\hat{v}(s)} \leq v(x) \sup_{s \in S} \frac{|\hat{Q}(s)|}{\hat{v}(s)} < \infty$$

holds for all $x \in X$ where the last inequality follows from $\hat{Q} \in \mathcal{B}_1(S)$ and $v(x)$ is finite for all $x \in X$. Therefore, we have $|(TJ)(x)| < \infty \forall x \in X$. In addition, by dividing $v(x)$ on both sides, we have

$$\frac{|(TJ)(x)|}{v(x)} \leq \sup_{s \in S} \frac{|\hat{Q}(s)|}{\hat{v}(s)}, \forall x \in X.$$

Therefore, we have

$$\sup_{x \in X} \frac{|(TJ)(x)|}{v(x)} \leq \sup_{s \in S} \frac{|\hat{Q}(s)|}{\hat{v}(s)} < \infty,$$

viz., $TJ \in \mathcal{B}(X)$.

On the other hand, $T_\mu J \in \mathcal{B}(X)$ and $TJ \in \mathcal{B}(X) \forall J \in \mathcal{B}(X)$, does not imply that $\hat{Q} \in \mathcal{B}_1(S) \forall J \in \mathcal{B}(X)$ where $\hat{Q}(s) = Q(x, u) = H(x, u, J)$. One example is found below.

Example 2.5.2. Consider $X = \{x\}$ and $U(x) = \mathbb{N}$. In addition, we define $v(x) = 1$ and $H(\cdot, \cdot, \cdot)$ as

$$H(x, u, J) = u + \alpha J(x)$$

where $\alpha \in (0, 1)$. Then we have $T_\mu J \in \mathcal{B}(X) \forall \mu \in \mathcal{M}$ and $TJ \in \mathcal{B}(X)$, but $\hat{Q} \notin \mathcal{B}_1(S)$.

P. 90

Here 'all' means for all $V, \tilde{V} \in \mathcal{B}(X)$ and all Q, \tilde{Q} such that $\hat{Q}, \hat{\tilde{Q}} \in \mathcal{B}_1(S)$ where $\hat{Q}(s) = Q(x, u)$ and $\hat{\tilde{Q}}(s) = \tilde{Q}(x, u)$ for $s = (x, u)$.

P. 91

Due to the definition of F_μ , we see that given V, Q , the function $F_\mu(V, Q)(\cdot, \cdot)$ is a function of (x, u) . Therefore, the norm here refers to the norm defined on $Q(\cdot, \cdot)$.

P. 91

Assume that $\hat{Q}(s) = Q(x, u) = H(x, u, J)$ is in the space $\mathcal{B}_1(S) \forall J \in \mathcal{B}(X)$, rather than assuming $T_\mu J \in \mathcal{B}(X)$ and $TJ \in \mathcal{B}(X) \forall J \in \mathcal{B}(X)$. Here we would like to show that this step follows from $\|T_\mu J - T_\mu J'\| \leq \alpha \|J - J'\|$

$\forall \mu \in \mathcal{M}$. Follow the notations defined above, we denote as S_μ the set $\cup_{x \in X} \{(x, \mu(x))\}$. Then

$$\begin{aligned} \|T_\mu J - T_\mu J'\| &= \sup_{x \in X} \frac{|H(x, \mu(x), J) - H(x, \mu(x), J')|}{v(x)} \\ &= \sup_{(x, u) \in S_\mu} \frac{|H(x, u, J) - H(x, u, J')|}{v(x)} \\ &= \sup_{s \in S_\mu} \frac{|\hat{Q}(s) - \hat{Q}'(s)|}{\hat{v}(s)} \end{aligned}$$

where $\hat{Q}(s) = H(x, u, J)$ and $\hat{Q}'(s) = H(x, u, J')$. Since $\|T_\mu J - T_\mu J'\| \leq \alpha \|J - J'\| \forall \mu \in \mathcal{M}$, we have

$$\sup_{s \in S_\mu} \frac{|\hat{Q}(s) - \hat{Q}'(s)|}{\hat{v}(s)} \leq \alpha \|J - J'\|, \forall \mu \in \mathcal{M},$$

which is to say, given $J, J' \in \mathcal{B}(X)$, $\forall \mu \in \mathcal{M}$, $\alpha \|J - J'\|$ is an upper bound of the set $\{|\hat{Q}(s) - \hat{Q}'(s)|/\hat{v}(s)\}_{s \in S_\mu}$. Consequently $\alpha \|J - J'\|$ is an upper bound of the set $\{|\hat{Q}(s) - \hat{Q}'(s)|/\hat{v}(s)\}_{s \in \cup_{\mu \in \mathcal{M}} S_\mu}$. Since we have

$$S \subseteq \bigcup_{\mu \in \mathcal{M}} S_\mu, \bigcup_{\mu \in \mathcal{M}} S_\mu \subseteq S \implies S = \bigcup_{\mu \in \mathcal{M}} S_\mu$$

where S contains all the feasible state control pairs (x, u) . One may verify the above relation by definitions of \mathcal{M} , S , S_μ . Therefore, we have $\alpha \|J - J'\|$ as an upper bound of the set $\{|\hat{Q}(s) - \hat{Q}'(s)|/\hat{v}(s)\}_{s \in S}$, viz.,

$$\|\hat{Q} - \hat{Q}'\| = \sup_{s \in S} \frac{|\hat{Q}(s) - \hat{Q}'(s)|}{\hat{v}(s)} \leq \alpha \|J - J'\|.$$

P. 91

Here 'all' means for all Q, \tilde{Q} such that $\hat{Q}, \hat{\tilde{Q}} \in \mathcal{B}_1(S)$ where $\hat{Q}(s) = Q(x, u)$ and $\hat{\tilde{Q}}(s) = \tilde{Q}(x, u)$ for $s = (x, u)$. For all such Q , we have $MQ \in \mathcal{B}(X)$. The proof is entirely similar to the one given for $\hat{Q} \in \mathcal{B}_1(S) \forall J \in \mathcal{B}(X)$ implying $TJ \in \mathcal{B}(X) \forall J \in \mathcal{B}(X)$. Here we repeat the arguments.

$$(MQ)(x) = \inf_{u \in U(x)} Q(x, u) = \inf_{s \in S_x} \hat{Q}(s) = v(x) \inf_{s \in S_x} \frac{\hat{Q}(s)}{\hat{v}(s)}.$$

Since

$$-\frac{|\hat{Q}(s)|}{\hat{v}(s)} \leq \frac{\hat{Q}(s)}{\hat{v}(s)} \leq \frac{|\hat{Q}(s)|}{\hat{v}(s)} \implies -\sup_{s \in S_x} \frac{|\hat{Q}(s)|}{\hat{v}(s)} \leq \inf_{s \in S_x} \frac{\hat{Q}(s)}{\hat{v}(s)} \leq \sup_{s \in S_x} \frac{|\hat{Q}(s)|}{\hat{v}(s)},$$

we have $\forall x \in X$

$$|(MQ)(x)| = \left| v(x) \inf_{s \in S_x} \frac{\hat{Q}(s)}{\hat{v}(s)} \right| \leq v(x) \sup_{s \in S_x} \frac{|\hat{Q}(s)|}{\hat{v}(s)} \leq v(x) \sup_{s \in S} \frac{|\hat{Q}(s)|}{\hat{v}(s)} < \infty,$$

namely $|MQ(x)| < \infty$ for all x . Dividing both sides with $v(x)$ and taking supremum over $x \in X$, we see that $MQ \in \mathcal{B}(X)$.

P. 91

In addition to this, it also follows from the relation

$$\|Q_\mu - \tilde{Q}_\mu\| \leq \|Q - \tilde{Q}\|,$$

and this can be verified as follows. Since

$$\begin{aligned} \|Q_\mu - \tilde{Q}_\mu\| &= \sup_{x \in X} \frac{|Q_\mu(x) - \tilde{Q}_\mu(x)|}{v(x)} = \sup_{s \in S_\mu} \frac{|\hat{Q}(s) - \hat{\tilde{Q}}(s)|}{\hat{v}(s)}, \\ \|Q - \tilde{Q}\| &= \sup_{x \in X, u \in U(x)} \frac{|Q(x, u) - \tilde{Q}(x, u)|}{v(x)} = \sup_{s \in S} \frac{|\hat{Q}(s) - \hat{\tilde{Q}}(s)|}{\hat{v}(s)}, \end{aligned}$$

the inequality is obtained since $S_\mu \subseteq S$.

P. 93

This update on J^t can be viewed as an combination of an step of local policy evaluation of Q on the old policy μ^t , and picking out the term of the evaluated Q whose control input corresponds to the new control μ^{t+1} . To see this, we have $\forall x \in X_\ell$

$$\begin{aligned} J^{t+1}(x) &= \min_{u \in U(x)} H(x, u, \min\{V^t, J^t\}) \\ &= H(x, \mu^{t+1}(x), \min\{V^t, J^t\}) \\ &= H(x, \mu^{t+1}(x), \min\{V^t, Q_{\mu^t}^t\}). \end{aligned}$$

Therefore, this step is performing policy evaluation of Q on the old policy (together with picking out the term of interests), and policy improvement on V, μ simultaneously.

3

Semicontractive Models

3.1 Pathologies of Noncontractive DP Models

3.1.1 Deterministic Shortest Path Problems

None.

3.1.2 Stochastic Shortest Path Problems

None.

3.1.3 The Blackmailer's Dilemma

None.

3.1.4 Linear-Quadratic Problems

P. 122

To see this, note that the sequence generated by the iteration is given as

$$p, r^2 + p(\gamma + r)^2, \dots, r^2 + \dots + (\gamma + r)^{2n-2}r^2 + p(\gamma + r)^{2n}, \dots$$

Therefore, the limit is $\frac{r^2}{1-(\gamma+r)^2}$ since $|\gamma + r| < 1$.

P. 122

Denote as p_0x^2 the cost function J_{μ^0} and as r_0x the control $\mu^0(x)$. Then we have $p_0 \geq \gamma^2 - 1$ and

$$\mu^1(x) = -\frac{p_0\gamma}{1+p_0}x,$$

as is shown in P. 120, [Abstract DP] 2nd Edition, which indeed is linear feedback and $r_1 = -\frac{p_0\gamma}{1+p_0}$. Then we have

$$|\gamma + r_1| = \frac{|\gamma|}{1+p_0} \leq \frac{|\gamma|}{1+\gamma^2-1} = \frac{1}{|\gamma|} < 1$$

where the first inequality follows from that $p_0 \geq \gamma^2 - 1$. Therefore, μ^1 is also stable linear control.

3.1.5 An Intuitive View of Semicontractive Analysis

P. 123

A direct consequence of this part of the assumption is that \hat{J} , defined point-wise by

$$\hat{J}(x) = \inf_{\mu \in \widehat{\mathcal{M}}} J_\mu(x),$$

can only takes values in $\mathbb{R} \cup \{-\infty\}$, viz., $\hat{J}(x) < \infty \forall x \in X$.

P. 124

Due to Eq. (3.8), we have $TJ_{\mu^k}(x) \in [J_{\mu^k}(x), J_{\mu^{k+1}}(x)] \subseteq \mathbb{R}$. Therefore, $\forall x \in X$, the sequence $\{TJ_{\mu^k}(x)\}_{k=0}^\infty$ is a real sequence. In addition, it is monotonically decreasing due to the monotonicity assumption of T and $J_{\mu^k} \geq J_{\mu^{k+1}}$. Therefore, $\forall x \in X$ $\{TJ_{\mu^k}(x)\}_{k=0}^\infty$ is convergent.

P. 125

To see that $\forall x \in X$, the sequence $\{T^k J(x)\}_{k=0}^\infty$ is convergent, we use the following arguments. We first have

$$T_\mu^k J \geq T^k J \geq T^k \hat{J} = \hat{J}, \forall \mu \in \widehat{\mathcal{M}}, k \geq 0. \quad (3.1)$$

Take limit inferior on both sides of Eq. (3.1) and since that $T^\mu J \rightarrow J_\mu$ is converging point-wise, we have

$$J_\mu(x) = \lim_{k \rightarrow \infty} (T_\mu^k J)(x) = \liminf_{k \rightarrow \infty} (T_\mu^k J)(x) \geq \liminf_{k \rightarrow \infty} (T^k J)(x) \geq \hat{J}(x).$$

Then we have $\forall x \in X$, it holds that

$$\hat{J}(x) = \inf_{\mu \in \widehat{\mathcal{M}}} J_\mu(x) \geq \liminf_{k \rightarrow \infty} (T^k J)(x) \geq \hat{J}(x) \implies \liminf_{k \rightarrow \infty} (T^k J)(x) = \hat{J}(x).$$

Similarly, we can get

$$\limsup_{k \rightarrow \infty} (T^k J)(x) \geq \hat{J}(x).$$

Therefore, we have $\forall x \in X$, the sequence $\{(T^k J)(x)\}_{k=0}^\infty$ is convergent.

3.2 Semicontractive Models and Regular Policies

3.2.1 S -Regular Policies

P. 128

3.2.2 Restricted Optimization over S -Regular Policies

P. 131

Lemma (P. 131). *Given nonempty set $S \subset \mathcal{E}(X)$, the relation of \mathcal{M}_S and \mathcal{W}_S being empty set is given as follows:*

$$\begin{aligned}\mathcal{M}_S \neq \emptyset &\implies \mathcal{W}_S \neq \emptyset; \\ \mathcal{W}_S \neq \emptyset &\implies (\mathcal{M}_S \neq \emptyset) \vee (J_\infty \in S);\end{aligned}$$

where J_∞ denotes the constant function that is equal to $+\infty \forall x \in X$.

Proof. For the first part, since $\mathcal{M}_S \neq \emptyset$, then $\exists \mu \in \mathcal{M}$ that is S -regular. Therefore, $J_\mu \in \mathcal{W}_S$ and $\mathcal{W}_S \neq \emptyset$. On the other hand, if $(\mathcal{M}_S = \emptyset) \wedge (J_\infty \notin S)$, we have $J_S^* = J_\infty$ and $\forall J \in S, J < J_S^*$ due to $J_\infty \notin S$, which indicates that $\mathcal{W}_S = \emptyset$. Taking logical not in above claim and we prove the second part. \square

Appendices

Appendix A

Notation and Mathematical Conventions

A.1 Set notion and conventions

None.

A.2 Functions

P. 324

Lemma (P. 324). *Let $\{s_{mn}\} \subset \mathbb{R}^*$ be an extended real-valued double sequence, which is monotonically nondecreasing separately for each index in the sense that*

$$s_{mn} \leq s_{(m+1)n}, \quad s_{mn} \leq s_{m(n+1)}, \quad \forall m, n = 0, 1, \dots,$$

then it holds that

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

Proof. If $\{s_{mn}\}$ is bounded above and $\{s_{mn}\} \cap \mathbb{R} \neq \emptyset$, then $\exists s_{MN} \in \mathbb{R}$. Since $\{s_{mn}\}$ is monotonically nondecreasing, then $s_{\ell k} \in \mathbb{R} \quad \forall \ell > M, k > N$. Then it is shown in Theorem 4.2, [Note 1] that the double sequence has real limit $\sup s_{mn}$. If $\{s_{mn}\}$ is bounded above and $\{s_{mn}\} \cap \mathbb{R} = \emptyset$, then $s_{mn} = -\infty \quad \forall m, n$, then the convergence of the double sequence follows. If $\{s_{mn}\}$ is unbounded above, then $\forall x \in \mathbb{R}, \exists s_{MN} > x$, since $\{s_{mn}\}$ nondecreasing, then $s_{\ell k} > x \quad \forall \ell > M, k > N$. Therefore, $\lim_{m, n \rightarrow \infty} s_{mn}$ exists and is ∞ according to definition. In conclusion, we prove that the double sequence converges and has a limit in \mathbb{R}^* .

In addition, $\forall m$, the sequence $\{s_{mn}\}_{n=0}^{\infty}$ is a monotone sequence in \mathbb{R}^* , then by [Lemma 3 P. 42], it is convergent. Then by the [Hand Note 1], we get

$$\lim_{m \rightarrow \infty} \left(\lim_{n \rightarrow \infty} s_{mn} \right) = \lim_{m, n \rightarrow \infty} s_{mn}.$$

Per definition of the limit of the double sequence, we have

$$\lim_{m,n \rightarrow \infty} s_{mn} = \lim_{m \rightarrow \infty} s_{mm},$$

which conclude the proof. \square

Appendix B

Contraction Mappings

B.1 Contraction mapping fixed point theorem

None.

B.2 Weighted sup-norm contractions

P. 329

The original claim is

$$\forall \varepsilon > 0, \exists K(\varepsilon) \in \mathbb{N}_+, \text{ such that } P(K(\varepsilon), \varepsilon)$$

where

$$P(K(\varepsilon), \varepsilon) = \left(\frac{|J_k(x) - J^*(x)|}{v(x)} \leq \varepsilon, \forall x \in X, k \geq K(\varepsilon) \right).$$

Therefore, the contrary is

$$\exists \varepsilon > 0, \forall K \in \mathbb{N}_+, \text{ such that } \neg P(K, \varepsilon)$$

where

$$\neg P(K, \varepsilon) = \left(\exists (x(K, \varepsilon) \in X, k(K, \varepsilon) \geq K), \frac{|J_{k(K, \varepsilon)}(x(K, \varepsilon)) - J^*(x(K, \varepsilon))|}{v(x(K, \varepsilon))} > \varepsilon \right).$$

Therefore, given ε , the sequence $\{x_{m_1}, x_{m_2}, \dots\}$ is constructed by induction: $x_{m_1} = x(1, \varepsilon)$ where $m_1 = k(1, \varepsilon)$; given x_{m_n} and m_n , $x_{m_{n+1}} = x(m_n + 1, \varepsilon)$ and $m_{n+1} = k(m_n + 1, \varepsilon)$.