DP2 READING GROUP

Longye Tian

Sydney – September 2024

Available at `https://github.com/longye-tian`

Definition

We define an **abstract dynamic program (ADP)** to be a pair $(V, \mathbb{T})$, where

1. $V = (V, \precsim)$: partially ordered set and
2. $\mathbb{T} = \{T_\sigma : \sigma \in \Sigma\}$: nonempty family of order-preserving self-maps on V.

In what follows,

- V is called the **value space**
- Each operator $T_\sigma \in \mathbb{T}$ is called a **policy operator**
- Σ is an arbitrary index set and elements of Σ is called **policies**.
- $v_\sigma \in V$ is a unique fixed of T_σ and call it the σ**-value function**

GREEDY POLICIES

Definition

Let $(V, \mathbb{T})$ be an ADP with policy set Σ. Given $v \in V$, we say that

$$\sigma \in \Sigma \text{ if } v\text{-}\textbf{greedy} \text{ if } T_\tau v \precsim T_\sigma v \text{ for all } \tau \in \Sigma$$

In other words, σ is v-greedy if and only if $T_\sigma v$ is a greatest element of $\{T_\tau v : T_\tau \in \mathbb{T}\}$

Remark

We let

$$V_G := \{v \in V : \textit{ at least one } v\textit{-greedy policy exists}\}$$

Definition

Let $(V, \mathbb{T})$ be an ADP with policy set Σ. We say that $v \in V$ satisfies the **Bellman Equation** if

$$v := \bigvee_{\sigma \in \Sigma} T_\sigma v \qquad (v \in V)$$

We define the **Bellman operator generated by the ADP** via

$$Tv := \bigvee_{\sigma \in \Sigma} T_\sigma v \qquad \text{whenever the supremum exists}$$

LEMMA 1.2.1.

Lemma (Properties of Bellman Operator)

1. *For $v \in V_G$, $T_\sigma v = Tv$ if and only if $\sigma \in \Sigma$ is v-greedy*

Proof.

$(\implies)$

$T_\sigma v = Tv \implies T_\sigma v = \bigvee_{\tau \in \Sigma} T_\tau v \implies T_\tau v \precsim T_\sigma v,\ \forall \tau \in \Sigma$ *by definition,* σ *is the* v*-greedy policy.* $v \in V_G$ *ensures existence.*

$(\impliedby)$

By definition, $T_\tau v \precsim T_\sigma v,\ \forall \tau \in \Sigma$. *Hence, there exists a greatest element hence supremum of* $\{T_\tau v : T_\tau \in \mathbb{T}\}$. *By definition, we have* $T_\sigma v = Tv$. $\square$

Lemma (Properties of Bellman Operator)

2. *For $v \in V_G$, we have $T_\sigma v \precsim Tv$ for all $\sigma \in \Sigma$.*

Proof.

As $v \in V_G$, there exists a v-greedy policy, denoted it as τ. Then, by part 1, we have by definition of v-greedy policy,

$$T_\sigma v \precsim T_\tau v = Tv$$

for all $\sigma \in \Sigma$. $\square$

Lemma (Properties of Bellman Operator)

3. *T is well-defined and order-preserving on V_G*

Proof.

For $v, w \in V_G$, there exists at least one v-greedy policy, denoted it as σ and let $v \precsim w$.

From part 1, we know $Tv = T_\sigma v$. Hence, it is well-defined.

From definition of ADP, we know T_σ is order-preserving, hence, we have,

$$Tv = T_\sigma v \precsim T_\sigma w \precsim Tw$$

last step is by part 2. Hence T is order-preserving on V_G. □

PROPERTIES OF ADP

We call $(V, \mathbb{T})$

- **well-posed** if each $T_\sigma \in \mathbb{T}$ has a unique fixed point in V
- **regular** if $V_G = V$, i.e., if a v-greedy policy exists for every $v \in V$.
- **bounded above** if there exists a $v \in V$ such that $T_\sigma v \precsim v$ for all $T_\sigma \in \mathbb{T}$
- **downward stable** if each $T_\sigma \in \mathbb{T}$ is downward stable on V
- **upward stable** if each $T_\sigma \in \mathbb{T}$ is upward stable on V
- **order stable** if each $T_\sigma \in \mathbb{T}$ is order stable on V
- **strongly order stable** if each $T_\sigma \in \mathbb{T}$ is strongly order stable on V
- **order continuous** if each $T_\sigma \in \mathbb{T}$ is order continuous on V

LEMMA 1.2.2

Lemma

If $(V, \mathbb{T})$ is order continuous and V is σ-chain complete. Then the Bellman operator T is order continuous.

Proof.

Let $v_n \uparrow v$. As T is order preserving, (Tv_n) is increasing.

Since V is σ-chain complete, $\bigvee_n Tv_n \in V$. We want to show $\bigvee_n Tv_n = Tv$

First, we have $Tv_n \leq Tv$ for all n by order-preserving, this implies $\bigvee_n Tv_n \leq Tv$. Hence, Tv is an upper bound.

Let w be an upper bound of Tv_n. Then we have $T_\sigma v_n \leq w$ for all $\sigma \in \Sigma$ and for all n. Then, taking the supremum over n, we have $T_\sigma v \leq w$ by order continuity of the ADP. Taking supremum over $\sigma \in \Sigma$, we have $Tv \leq w$. Hence Tv is the least upper bound. □

EXERCISE 1.2.3.

Let $(V, \mathbb{T})$ be an ADP and let V be σ-dedekind complete. Prove that if $(V, \mathbb{T})$ is order continuous and order stable, then $(V, \mathbb{T})$ is strongly order stable.

Proof.

Implore the Tarski-Kantorovich I (Theorem 1.1.9). $\square$

SUBSETS OF THE VALUE SPACE

Definition

Let $(V, \mathbb{T})$ be an ADP. We oftern refer to the following three subsets of the value space

- V_G := all $v \in V$ such that at least one v-greedy policy exists
- V_U := all $v \in V$ with $v \precsim Tv$
- V_Σ := all $v \in V$ such that $T_\sigma v = v$ for some $T_\sigma \in \mathbb{T}$

LEMMA 1.2.3

Lemma

Let $(V, \mathbb{T})$ be regular, well-posed and upward stable. If V_Σ has greatest element v^*, then $v \precsim v^*$ for all $v \in V_U$.

Proof.

Fix $v \in V_U$. By regularity, we let σ be the v-greedy policy, then

$$\underbrace{v \precsim Tv}_{v \in V_U} \quad \underbrace{= T_\sigma v}_{\text{Lemma 1.2.1}} \quad \underbrace{\implies v \le v_\sigma}_{\text{upward stable}} \quad \underbrace{\le v^*}_{\text{greatest element}}$$

$\square$

LEMMA 1.2.4.

Lemma

Let $(V, \mathbb{T})$ be an ADP. If $V_\Sigma \subset V_G$, then $V_\Sigma \subset V_U$

Proof.

Fix $v \in V_\Sigma \subset V_G$. Let v_σ be σ-value function and let τ be v_σ-greedy policy. By Lemma 1.2.1, we have

$$v_\sigma = T_\sigma v_\sigma \precsim T_\tau v_\sigma = T v_\sigma$$

Hence, $v_\sigma \in V_U$ and $V_\Sigma \subset V_U$. □

Remark

Lemma 1.2.4. implies that if $(V, \mathbb{T})$ is well-posed, then $V_U \neq \varnothing$

OPTIMALITY AND BELLMAN EQUATION

Definition

We say that a policy $\sigma \in \Sigma$ is **optimal** for $(V, \mathbb{T})$ if v_σ is a greatest element of V_Σ.

In other words, σ is optimal if it attains the highest possible lifetime value.

Definition

Suppose V_Σ has a greatest element v^*, which is called the **value function of the ADP**. We say that **Bellman's principle of optimality holds** if, for $\sigma \in \Sigma$,

$$\sigma \text{ is optimal} \iff \sigma \text{ is the } v^*\text{-greedy policy}$$

THREE FUNDAMENTAL ADP OPTIMALITY PROPERTIES

We say that the **fundamental ADP optimality properties holds** if

(B1) V_Σ has a greatest element v^*

(B2) v^* is the unique solution to the Bellman equation

(B3) Bellman's principle of optimality holds

Remark

(B1) *means that the ADP has a solution*

(B2) *characterize the solution*

(B3) *implies we can compute the solution by finding v^*-greedy policy*

- *(B1) and (B2) $\implies$ (B3)*

LEMMA 1.2.5.

Let $(V, \mathbb{T})$ be a well-posed ADP. If V_Σ has a greatest element v^*, the TFAE:

1. v^* satisfies the Bellman equation, i.e., $v^* = \bigvee_\tau T_\tau v^*$
2. Bellman's principle of optimality holds

Proof.

$((1) \Rightarrow (2), \Rightarrow)$:

$T_m v^* \precsim \bigvee_\tau T_\tau v^* = v^* = T_\sigma v^* \, \forall T_m \in \mathbb{T}$

$((1) \Rightarrow (2), \Leftarrow)$:

$T_m v^* \precsim T_\sigma v^* \, \forall T_m \in \mathbb{T} \Rightarrow \bigvee_\tau T_\tau v^* \precsim T_\sigma v^* \Rightarrow v^* \precsim T_\sigma v^* \precsim \bigvee_\tau T_\tau v^* = v^*$

$((1) \Leftarrow (2))$:

Let σ be v^*-greedy policy. Then, $T_\tau v^* \precsim T_\sigma v^* \, \forall T_\tau \in \mathbb{T}$

Hence, v^* is an upper bound of $\{T_\tau v^*\}_{T_\tau \in \mathbb{T}}$. Let $\tilde{v}$ be any upper bound.

Then, $v^* = T_\sigma v^* \precsim \tilde{v}$. Hence $v^* = \bigvee_\tau T_\tau v^*$ □

LEMMA 1.2.6.

Let $(V, \mathbb{T})$ be a well-posed ADP. Let V_Σ have the greatest element v^*. If v^* is unique fixed point of T in V, then $\sigma \in \Sigma$ is optimal if and only if $Tv_\sigma = v_\sigma$.

Proof.

$(\Rightarrow)$

$v_\sigma = v^* \Rightarrow Tv_\sigma = Tv^* = v^* = v_\sigma$

$(\Leftarrow)$

$Tv_\sigma = v_\sigma \Rightarrow v_\sigma = v^*$ □

ALGORITHMS

HPI, OPI, VFI

Definition

Let $(V, \mathbb{T})$ be a well-posed ADP with Bellman operator T.

The **Howard Policy Operator** corresponding to $(V, \mathbb{T})$ via

$$H : V_G \to V_\Sigma, \qquad Hv = v_\sigma, \qquad \sigma \text{ is } v\text{-greedy}$$

For each $m \in \mathbb{N}$, the **optimistic policy operator**

$$W_m : V_G \to V, \qquad W_m v := T_\sigma^m v, \qquad \text{where } \sigma \text{ is } v\text{-greedy}$$

So that these maps are well-defined, we always select the same v-greedy policy when applying each to v.

CONVERGENCE

Let $(V, \mathbb{T})$ be a well-posed ADP. Suppose that the fundamental ADP optimality properties hold. Let v^* denote the value function. We say that

- **VFI converges** if $T^n v \uparrow v^*$ for all $v \in V_U$
- **OPI converges** if $W_m^n v \uparrow v^*$ for all $v \in V_U$ and all $m \in \mathbb{N}$
- **HPI converges** if $H^n v \uparrow v^*$ for all $v \in V_U$
- If, for all $v \in V_U$, there exists $n \in \mathbb{N}$ with $H^n v = v^*$, we say that **HPI converges in finitely many steps**.

EXERCISE 1.2.4

Prove that convergence of OPI implies convergence of VFI.

Proof.

Let $m = 1$. We have $W_1 v = T_\sigma v = Tv =: v_1$, $W_1^2 v = W_1 v_1 = Tv_1 = T^2 v$. By induction, we get $W_1^n v = T^n v$. Hence, we get OPI convergence implies VFI convergence. $\square$

LEMMA 1.2.7.

If $(V, \mathbb{T})$ is regular and well-posed, then the following statement hold.

L1 If $v \in V$ with $Hv = v$, then $Tv = v$

Proof.

$Hv = v_\sigma = v \Rightarrow Tv = T_\sigma v = T_\sigma v_\sigma = v_\sigma = v$ □

L2 The operators T, W_m, H all maps V_U to itself

Proof.

$Tv \geq v$ and by T is order-preserving on V_G and by regularity, we get $T(Tv) \geq Tv$.

$W_m v = T_\sigma^m v = T_\sigma^{m-1} Tv \geq v$ by $v \in V_U$ and T_σ^{m-1} is order-preserving and by W_m is order-preserving, we get the claim.

$Hv = v_\sigma$. By regularity, we get $V_\Sigma \subset V_G \implies V_\Sigma \in V_U$. □

LEMMA 1.2.7. CONTINUE

If $(V, \mathbb{T})$ is regular and well-posed, then the following statement hold.

L3 The operator W_m is order preserving on V_U and

$$v \in V_U \implies Tv \lesssim W_m v \lesssim T^m$$

Proof.

Order preserving is from order-preserving of T_σ.

$W_m v = T_\sigma^m v = T_\sigma^{m-1} Tv = Wm - 1Tv$. The by W_m is order preserving, we have $W_m v = W_{m-1} Tv \geq W_{m-1} v$. Iteratively, we get $W_m \geq Tv$.

For the second inequality, we have

$$W_m v = T_\sigma^m v \lesssim T_\sigma^{m-1} Tv \lesssim T_\sigma^{m-2} T^2 v$$

Iteratively, we get the inequality. □

LEMMA 1.2.8.

If $(V, \mathbb{T})$ is well-posed, regular and upward stable, then for every $v \in V_U$,

$$v \precsim T^n v \precsim W_m^n v \precsim H^n v$$

for all $n \in \mathbb{N}$. And the VFI sequence $(T^n v)$, OPI sequence $(W_m^n v)$ and HPI sequence $(H^n v)$ are all increasing.

Corollary

Let $(V, \mathbb{T})$ be regular and upward stable. If the fundamental ADP optimality properties hold, then convergence of VFI implies convergence of OPI and HPI.

OPTIMALITY VIA ORDER STABILITY

CONDITIONS FOR ADP FUNDAMENTAL PROPERTIES HOLD

Theorem

Let $(V, \mathbb{T})$ be well-posed, downward stable and T has a fixed point in V_G, then

1. *the fixed point of T is the greatest element of V_Σ*
2. *T has no other fixed point in V_G*
3. *σ is optimal if and only if σ is v-greedy.*

COROLLARY 1.2.11

Let $(V, \mathbb{T})$ be well-posed, regular and downward stable. If T has a fixed point in V, then T has exactly one fixed point in V and moreover, the fundamental ADP optimality properties hold.

FINITE CASE - THEOREM 1.2.12

Let $(V, \mathbb{T})$ be well-posed, regular and order stable. If $\mathbb{T}$ is finite, then

1. the fundamental ADP optimality properties hold
2. HPI converges in finitely many steps