DSE 2023 Summer School Lausanne

Lecture 2: Advances in DP: Theory and Algorithms

John Stachurski

2023

Topics

Handling more general dynamic programs

- recursive preferences, quantile preferences, etc.
- recursive preferences + state-dependent discounting, etc.

Abstraction to simplify proofs

- clarifies optimality proofs
- clarifies relationships between DPs

Motivation

Consider

$$\max \sum_{t \geqslant 0} \beta^t u(C_t)$$

subject to

$$W_{t+1} = R(W_t - C_t) \quad \text{and} \quad 0 \leqslant C_t \leqslant W_t$$

Standard approach: set up the **Bellman operator**

$$(Tv)(w) = \max_{0 \leqslant c \leqslant w} \{u(c) + \beta v(R(w - c))\}$$

Value function iteration (VFI)

Under some conditions,

1. T is a contraction mapping
2. the unique fixed point of T is the value function v^*
3. v^* can be approximated via $v^* = \lim_{k\to\infty} T^k v$ for some v
4. optimal consumption at wealth w can be found by solving

$$c^* \in \underset{0\leqslant c\leqslant w}{\operatorname{argmax}} \{u(c) + \beta v^*(R(w-c))\}$$

Howard policy iteration

Alternatively, we can use Howard policy iteration (HPI)

A **feasible policy** is a map $\sigma \colon \mathbb{R}_+ \to \mathbb{R}_+$ with

$$0 \leqslant \sigma(w) \leqslant w \quad \text{for all} \quad w \in \mathbb{R}_+$$

- given current wealth w, choose consumption $c = \sigma(w)$
- $\Sigma :=$ all feasible policies

A feasible policy σ is called v-**greedy** if

$$\sigma(w) \in \operatorname*{argmax}_{0 \leqslant c \leqslant w} \{u(c) + \beta v(R(w-c))\}$$

Algorithm 1: Howard policy iteration

input $\sigma_0 \in \Sigma$, set $k \leftarrow 0$ and $\varepsilon \leftarrow 1$

while $\varepsilon > 0$ **do**

$v_k \leftarrow$ the lifetime value of σ_k

$\sigma_{k+1} \leftarrow$ a v_k-greedy policy

$\varepsilon \leftarrow \mathbb{1}\{\sigma_k \neq \sigma_{k+1}\}$

$k \leftarrow k + 1$

end

return σ_k

Optimistic policy iteration

The lifetime value v_σ of policy σ is the unique v that solves

$$v(w) = u(\sigma(w)) + \beta v(R(w - \sigma(w)))$$

To compute it we introduce the **policy operator**

$$(T_\sigma\, v)(w) = u(\sigma(w)) + \beta v(R(w - \sigma(w)))$$

Facts:

1. v_σ is the unique fixed point of T_σ
2. $T_\sigma^k v \to v_\sigma$ as $k \to \infty$ for all reasonable v

Algorithm 2: OPI

input v_0, an initial guess of v^*

input τ, a tolerance level for error

input $m \in \mathbb{N}$, a step size

$k \leftarrow 0$

$\varepsilon \leftarrow +\infty$

while $\varepsilon > \tau$ **do**

$\sigma_k \leftarrow$ a v_k-greedy policy

$v_{k+1} \leftarrow T_{\sigma_k}^m v_k$

$\varepsilon \leftarrow \|v_k - v_{k+1}\|_\infty$

$k \leftarrow k + 1$

end

return σ_k

State-dependent discounting

What happens if we introduce state-dependent discounting?

$$(Tv)(w,z) = \max_{0 \leqslant c \leqslant w} \left\{ u(c) + \beta(z) \sum_{z'} v(R(w-c), z') Q(z, z') \right\}$$

T is no longer a one-step contraction

However, as discussed in Lecture 1,

- Standard optimality results hold
- OPI, VFI, HPI all converge

More complications

What happens if we switch to the **expected value function**

$$g(w,z,c) := \sum_{z'} v(R(w-c),z')Q(z,z')$$

with "Bellman operator"

$$(Rg)(w,z,c) =$$

$$\sum_{z'} \max_{0 \leqslant c' \leqslant R(w-c)} \left\{ u(c') + \beta(z') g(R(w-c),z',c') \right\} Q(z,z')$$

Does R have the same properties as T?

What are the equivalent algorithms and do they converge?

And what happens if we introduce **Epstein–Zin preferences**?

$$(Tv)(w,z) =$$

$$\max_{0 \leqslant c \leqslant w} \left\{ c^\alpha + \beta(z) \left[\sum_{z'} v(R(w-c), z')^\gamma Q(z, z') \right]^{\alpha/\gamma} \right\}^{1/\alpha}$$

- Is T still a contraction?
- Are the previous optimality results still valid?
- Do VFI, OPI, HPI converge?

Or **risk-sensitive preferences**?

$$(Tv)(w,z) =$$

$$\max_{0 \leqslant c \leqslant w} \left\{ u(c) + \frac{\beta(z)}{\theta} \ln \left[\sum_{z'} \mathrm{e}^{\theta v(R(w-c),z')} Q(z,z') \right] \right\}$$

- Is T still a contraction?
- Are the previous optimality results still valid?
- Do VFI, OPI, HPI converge?

What about if we want to handle

- ambiguity?
- expected value VFI in an Epstein–Zin framework?
- expected value VFI + ambiguity + state-dependent discounting?
- integrated value functions in a risk-sensitive framework in continuous time?
- Q-learning?

Is there any unifying theory?

Or are all these problems too diverse?

Abstraction Level 1: RDPs

1. Construct a DP framework based on an abstraction of Bellman's equation
2. State optimality results in this framework
3. Connect with applications

Builds on work by

- Eric Denardo
- Dimitri Bertsekas
- Takashi Kamihigashi

Recursive Decision Problems

We begin with a generic version of Bellman's equation:

$$v(x) = \max_{a \in \Gamma(x)} B(x, a, v)$$

- $x \in$ a finite set X (the **state space**)
- $a \in$ a finite set A (the **action space**)
- $B(x, a, v) =$ total lifetime rewards
  - contingent on current state-action pair (x, a) and
  - using v to evaluate future states

Definition

A **recursive decision process** (RDP) is a triple (Γ, V, B), where

◦ Γ is a nonempty correspondence from X to A

called the **feasible correspondence**

Generates:

- the **feasible state-action pairs**

$$\mathsf{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\}$$

- the set of **feasible policies**

$$\Sigma := \{\sigma \in \mathsf{A}^{\mathsf{X}} : \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\}$$

- V is a sublattice of $\mathbb{R}^{\mathsf{X}}$ called the **value space**

  - A set of candidates for the value function

(Sublattice $\iff$ $f \vee g \in V$ and $f \wedge g \in V$ when $f, g \in V$)

- B is a map from $\mathsf{G} \times V$ to $\mathbb{R}$ called the **value aggregator**

It satisfies

1. **Monotonicity**:

$$v, w \in V \text{ and } v \leqslant w \implies B(x, a, v) \leqslant B(x, a, w)$$

for all $(x, a) \in \mathsf{G}$

2. **Consistency**:

$$w(x) := B(x, \sigma(x), v) \text{ is in } V \text{ whenever } \sigma \in \Sigma \text{ and } v \in V$$

Example. An arbitrary Markov Decision Process with Bellman equation

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{x'} v(x') P(x,a,x') \right\} \tag{1}$$

can be framed as an RDP

Take Γ as given, set $V = \mathbb{R}^{\mathsf{X}}$, and

$$B(x,a,v) = r(x,a) + \beta \sum_{x'} v(x') P(x,a,x')$$

- monotonicity and consistency conditions are trivial to check
- recover (1) via $v(x) = \max_{a \in \Gamma(x)} B(x,a,v)$

Example. Consider an **optimal stopping** problem with

$$v(x) = \max \left\{ e(x), c(x) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, x') \right\}$$

Let $V = \mathbb{R}^{\mathsf{X}}$

If $\Gamma(x) = \{0, 1\}$ and

$$B(x, a, v) = ae(x) + (1 - a) \left[c(x) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, x') \right]$$

then (Γ, V, B) is an RDP with the same Bellman equation

Example. Consider an MDP with **state-dependent discounting**, so that

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x'} v(x') \beta(x,a,x') P(x,a,x') \right\}$$

Let $V = \mathbb{R}^{\mathsf{X}}$ and

$$B(x,a,v) = r(x,a) + \sum_{x'} v(x') \beta(x,a,x') P(x,a,x')$$

Now (Γ, V, B) is an RDP with the same Bellman equation

Example. Consider a modified MDP with **risk-sensitive preferences**, so that

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \frac{1}{\theta} \ln \left(\sum_{x'} \exp(\theta v(x')) P(x,a,x') \right) \right\}$$

for nonzero θ

With $V = \mathbb{R}^{\mathsf{X}}$ and

$$B(x,a,v) = r(x,a) + \beta \frac{1}{\theta} \ln \left(\sum_{x'} \exp(\theta v(x')) P(x,a,x') \right)$$

we obtain an RDP with the same Bellman equation

Example. Consider a modified MDP with **quantile preferences**, so that

$$v(x) = \max_{a \in \Gamma(x)} \{r(x,a) + \beta(R_\tau^a v)(x)\}$$

where

$$(R_\tau^a v)(x) := \tau\text{-th quantile of } v(X') \text{ when } X' \sim P(x,a,\cdot)$$

With $V = \mathbb{R}^{\mathsf{X}}$ and

$$B(x,a,v) = r(x,a) + \beta(R_\tau^a v)(x)$$

we obtain an RDP with the same Bellman equation

Example. Consider a modified MDP with **Epstein–Zin preferences**, so that

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \left(\sum_{x'} v(x')^\gamma P(x,a,x') \right)^{\alpha/\gamma} \right\}^{1/\alpha}$$

for nonzero α, γ

With $V =$ the strictly positive functions in $\mathbb{R}^{\mathsf{X}}$ and

$$B(x,a,v) = \left\{ r(x,a) + \beta \left(\sum_{x'} v(x')^\gamma P(x,a,x') \right)^{\alpha/\gamma} \right\}^{1/\alpha}$$

we obtain an RDP with the same Bellman equation

Example. Consider a **shortest path problem** on graph $\mathscr{G} = (\mathsf{X}, E)$

- $c(x, x') =$ cost of traversing edge $(x, x') \in E$
- the direct successors of x denoted by

$$\mathscr{O}(x) := \{x' \in \mathsf{X} : (x, x') \in E\}$$

Aim: find the minimum cost path from x to a specified vertex d

No discounting (so cannot use MDP theory)

The Bellman equation is

$$v(x) = \min_{x' \in \mathscr{O}(x)} \{c(x, x') + v(x')\}$$

Let $V = \mathbb{R}^{\mathsf{X}}$

Let $\Gamma(x) = \mathscr{O}(x)$ and

$$B(x, x', v) = c(x, x') + v(x')$$

This is an RDP with the same Bellman equation

Policies

Consider an arbitrary RDP (Γ, V, B)

A **feasible policy** is a

$$\sigma \in \mathsf{A}^{\mathsf{X}} \text{ such that } \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}$$

- respond to state X_t with action $A_t := \sigma(X_t)$ at **all** $t \geqslant 0$
- $\Sigma :=$ the set of all feasible policies

Policy Operators

Fix $\sigma \in \Sigma$

The corresponding **policy operator** T_σ is defined at $v \in V$ by

$$(T_\sigma\, v)(x) = B(x, \sigma(x), v) \qquad (x \in \mathsf{X})$$

Lemma. T_σ is an order-preserving self-map on V

Proof: Immediate from monotonicity and consistency

Example. The EZ policy operator is

$$(T_\sigma\, v)(x) = \left\{ r(x, \sigma(x)) + \beta \left(\sum_{x'} v(x')^\gamma P(x, \sigma(x), x') \right)^{\alpha/\gamma} \right\}^{1/\alpha}$$

Optimality

To define optimality for RDPs, we use the natural generalizations...

Lifetime value

Let $\mathscr{R} := (\Gamma, V, B)$ be an RDP and let σ be any policy

Suppose T_σ has a unique fixed point in V

We denote this function by v_σ and call it the **σ-value function**

We interpret this function as the lifetime value of following σ

We call $\mathscr{R}$ **well-posed** if T_σ has a unique fixed point in V for all $\sigma \in \Sigma$

Example. Let $\mathscr{R}$ be the RDP generated by an MDP

Recall that

$$T_\sigma v = r_\sigma + \beta P_\sigma v$$

This operator has the unique fixed point

$$v_\sigma = (I - \beta P_\sigma)^{-1} r_\sigma$$

- Hence $\mathscr{R}$ is well-posed
- $v_\sigma(x) = \mathbb{E}_x \sum_{t \geqslant 0} \beta^t r(X_t, \sigma(X_t)) =$ lifetime value

Example. For the Epstein–Zin RDP,

$$(T_\sigma v)(x) = \left\{ r(x, \sigma(x)) + \beta \left[\sum_{x' \in \mathsf{X}} v(x')^\gamma P(x, \sigma(x), x') \right]^{\alpha/\gamma} \right\}^{1/\alpha}$$

and

$$V := \text{the strictly positive functions in } \mathbb{R}^{\mathsf{X}}$$

- Is this RDP well-posed?

Greedy Policies

Fix $v \in \mathbb{R}^{\mathsf{X}}$

A policy σ is called v-**greedy** if

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} B(x, a, v)$$

for all $x \in \mathsf{X}$

Note: at least one v-greedy policy exists in Σ

The Bellman Operator

The **Bellman operator** is the self-map on $\mathbb{R}^X$ defined by

$$(Tv)(x) = \max_{a \in \Gamma(x)} B(x, a, v)$$

Key idea:

$$Tv = v \quad \iff \quad v \text{ satisfies the Bellman equation}$$

Optimality

Let $\mathscr{R}$ be a well-posed RDP

The **value function** is defined by $v^* = \bigvee v_\sigma$

More explicitly,

$$v^*(x) := \max_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in \mathsf{X})$$

$$= \text{max lifetime value from state } x$$

A policy $\sigma \in \Sigma$ is called **optimal** if

$$v_\sigma = v^*$$

Howard policy iteration for RDPs

input $\sigma_0 \in \Sigma$, an initial guess of σ^*
$k \leftarrow 0$
$\varepsilon \leftarrow 1$
while $\varepsilon > 0$ **do**
    $v_k \leftarrow$ the unique fixed point of T_{σ_k}
    $\sigma_{k+1} \leftarrow$ a v_k greedy policy
    $\varepsilon \leftarrow \mathbb{1}\{\sigma_k \neq \sigma_{k+1}\}$
    $k \leftarrow k + 1$
end
return σ_k

Let $\mathscr{R}$ be an RDP

Key question:

What assumptions to we need for optimality?

Obviously $\mathscr{R}$ must be well-posed

- each T_σ has a unique fixed point in V

This is the minimum requirement

What else?

Stability

Let $\mathscr{R}$ be an RDP

We call $\mathscr{R}$ **globally stable** if, for all $\sigma \in \Sigma$, the operator T_σ is globally stable on V

Meaning

1. T_σ has a unique fixed point in V and
2. $\lim_{k\to\infty} T_\sigma^k v = v_\sigma$ for all $v \in V$

Let $\mathscr{R}$ be a well-posed RDP with value function v^*

Theorem. If $\mathscr{R}$ is stable, then

1. v^* is the unique solution to the Bellman equation in $\mathbb{R}^{\mathsf{X}}$
2. A feasible policy is optimal if and only it is v^*-greedy
3. At least one optimal policy exists
4. Howard policy iteration returns an optimal policy in finitely many steps

Types of RDPs

The optimality properties require stability

We can check this directly

We can also

1. identify classes of RDPs that are stable
2. show that a given application belongs to one of these classes

Let's discuss the classification approach

Below $\mathscr{R} = (\Gamma, V, B)$ is a fixed RDP

Contracting RDPs

We call $\mathscr{R}$ **contracting** if $\exists\ \beta < 1$ such that

$$|B(x, a, v) - B(x, a, w)| \leqslant \beta \|v - w\|_\infty$$

for all $(x, a) \in$ G and $v, w \in V$

Thm. If $\mathscr{R}$ is contracting and V is closed, then $\mathscr{R}$ is stable

Proof: Easy to show that each T_σ is a contraction on V

(Main idea dates back to Denardo 1967)

Eventually Contracting RDPs

We call $\mathscr{R}$ **eventually contracting** if there is an $L \geqslant 0$ such that $\rho(L) < 1$ and

$$|B(x, a, v) - B(x, a, w)| \leqslant \sum_{x'} |v(x') - w(x')| L(x, x')$$

for all $(x, a) \in \mathsf{G}$ and $v, w \in V$

Thm. If $\mathscr{R}$ is eventually contracting and V is closed, then $\mathscr{R}$ is stable

Proof: See the book

Concave RDPs

We call $\mathscr{R}$ **concave** if

1. $V = [v_1, v_2]$
2. $B(x, a, v_1) > v_1(x)$ for all $(x, a) \in \mathsf{G}$ and
3. $v \mapsto B(x, a, v)$ is concave for all $(x, a) \in \mathsf{G}$

Thm. If $\mathscr{R}$ is concave, then $\mathscr{R}$ is stable

Proof: See the book

Application: job search with quantile preferences

Set up:

- wage offer process $(W_t)_{t \geqslant 0}$ is P-Markov on finite set W
- discount factor $\beta \in (0, 1)$

The Bellman equation is

$$v(w) = \max \left\{ \frac{w}{1-\beta}, \, c + \beta (R_\tau v)(w) \right\}$$

Here

$$(R_\tau v)(w) := \tau\text{-th quantile of } v(W') \text{ when } W' \sim P(w, \cdot)$$

This problem studied in

- de Castro and Galvao (2019)
- de Castro, Galvao and Nunes (2022)
- de Castro and Galvao (2022)

We can embed the into the RDP framework by taking

- $\Gamma(w) = \{0, 1\}$
- $V = \mathbb{R}_+^{\mathsf{W}}$
- B given by

$$B(w, a, v) = a\frac{w}{1-\beta} + (1-a)[c + \beta(R_\tau v)(w)]$$

Easy to check that $\mathscr{R} := (\Gamma, V, B)$ is an RDP with Bellman equation

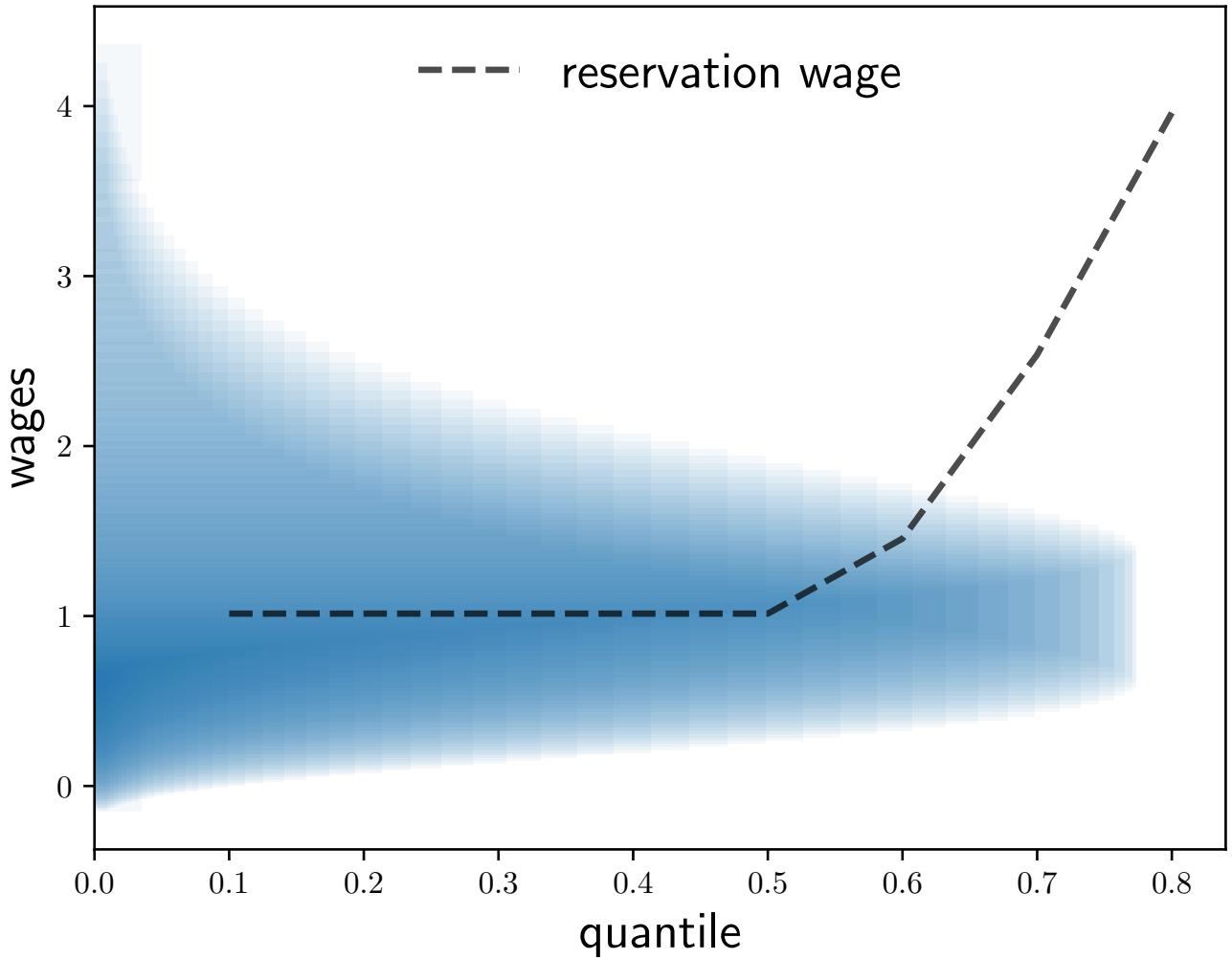
$$v(w) = \max\left\{\frac{w}{1-\beta},\, c + \beta(R_\tau v)(w)\right\}$$

Proposition. $\mathscr{R}$ is a contracting RDP

- Proof: See the text

Since V is closed, $\mathscr{R}$ is stable

Hence all optimality properties apply


reservation wage
wages
quantile
0
1
2
3
4
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

Abstraction Level 2: ADPs

We define an **abstract dynamic program** (**ADP**) to be a pair

$$\mathscr{A} = (V, \{T_\sigma\}_{\sigma \in \Sigma}), \qquad \text{where}$$

1. $V = (V, \preceq)$ is a partially ordered set and
2. $\{T_\sigma\}_{\sigma \in \Sigma}$ is a family of self-maps on V

Below,

- elements of Σ will be referred to as **policies**
- elements of $\{T_\sigma\}$ are called **policy operators**

If T_σ has a unique fixed point, then we

- denote it v_σ and call it the **σ-value function**
- understand v_σ as representing lifetime value of σ

Interpretation:

- V is a set of candidate value functions
- Σ is a set of feasible policies
- the lifetime value of $\sigma \in \Sigma$ is v_σ
- we seek a greatest element in $\{v_\sigma\}_{\sigma \in \Sigma}$

Example. Consider an RDP (Γ, V, B)

Let Σ be the set of feasible policies

Recall that, for each $\sigma \in \Sigma$, the policy operator T_σ is defined at $v \in V$ by

$$(T_\sigma\, v)(x) = B(x, \sigma(x), v)$$

The pair $(V, \{T_\sigma\})$ is an ADP

Benefits of ADP theory

- More abstraction means easier proofs
- Removing structure makes it easier to see connections
- Can handle a more diverse range of problems

Given $v \in V$, a policy σ in Σ is called **v-greedy** if

$$T_\sigma\, v \succeq T_\tau\, v \quad \text{for all } \tau \in \Sigma$$

Example. In the MDP example we have

$$(T_\sigma\, v)(x) = r(x, \sigma(x)) + \beta \sum_{x'} v(x') P(x, \sigma(x), x')$$

so σ is v-greedy iff

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{x'} v(x') P(x,a,x') \right\} \quad \text{for all } x \in \mathsf{X}$$

Bellman equation

Fix an ADP $\mathscr{A} = (V, \{T_\sigma\})$

We define the **Bellman operator** via

$$Tv := \bigvee_\sigma T_\sigma \, v$$

(if it exists)

Equivalently,

$$Tv = T_\sigma \, v \text{ when } \sigma \text{ is } v\text{-greedy}$$

We say that $v \in V$ satisfies the **Bellman equation** if $Tv = v$

Example. For the MDP,

$$(Tv)(x) = (T_\sigma\, v)(x) \text{ when } \sigma \text{ is } v\text{-greedy}$$

$$= \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{x'} v(x') P(x,a,x') \right\}$$

Hence the ADP Bellman equation is

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{x'} v(x') P(x,a,x') \right\}$$

And this is the same as the MDP Bellman equation

Properties

We say that $\mathscr{A} = (V, \{T_\sigma\})$ is

- **well-posed** if T_σ has one fixed point in V for each $\sigma \in \Sigma$
- **order stable** if (V, T_σ) is order stable for each $\sigma \in \Sigma$
- **max-stable** if $\mathscr{A}$ is order stable, each $v \in V$ has at least one greedy policy, and T has at least one fixed point in V

Note: order stability is a regularity property — see Ch 9

Let $\mathscr{A}$ be a well-posed ADP

A policy $\sigma \in \Sigma$ is called **optimal** for $\mathscr{A}$ if

$$v_\tau \preceq v_\sigma \text{ for all } \tau \in \Sigma$$

We set $v^* := \bigvee_\sigma v_\sigma$ and call v^* the **value function**

We define a self-map H on V via

$$H\, v = v_\sigma \quad \text{where} \quad \sigma \text{ is } v\text{-greedy}$$

Iterating with H is an abstract version of HPI

Max-Optimality

Theorem. If $\mathscr{A}$ is max-stable, then

1. v^* exists in V
2. v^* is the unique solution to the Bellman equation in V
3. a policy is optimal if and only if it is v^*-greedy
4. at least one optimal policy exists

If, in addition, Σ is finite, then HPI $\to v^*$ in finitely many steps

Subordinate ADPs

Let $\mathscr{A} := (V, \{T_\sigma\})$ and $\hat{\mathscr{A}} := (\hat{V}, \{\hat{T}_\sigma\})$ be ADPs

We say that $\hat{\mathscr{A}}$ is **subordinate** to $\mathscr{A}$ if $\exists$

1. an order-preserving map F from V onto $\hat{V}$ and
2. order-preserving maps $\{G_\sigma\}_{\sigma\in\Sigma}$ from $\hat{V}$ to V

such that

$$T_\sigma = G_\sigma \circ F \quad \text{and} \quad \hat{T}_\sigma = F \circ G_\sigma \qquad \text{for all } \sigma \in \Sigma$$

Let $G = \bigvee_\sigma G_\sigma$

Theorem. If

1. $\mathscr{A}$ is max-stable and
2. $\hat{\mathscr{A}}$ is subordinate to $\mathscr{A}$,

then $\hat{\mathscr{A}}$ is also max-stable and the Bellman operators are related by

$$T = G \circ F \quad \text{and} \quad \hat{T} = F \circ G$$

while the value functions are related by

$$v^* = G\,\hat{v}^* \quad \text{and} \quad \hat{v}^* = F\,v^*$$

Moreover,

1. if σ is optimal for $\mathscr{A}$, then σ is optimal for $\hat{\mathscr{A}}$, and
2. if $G_\sigma\,\hat{v}^* = G\,\hat{v}^*$, then σ is optimal for $\mathscr{A}$

Application

Consider an Epstein–Zin dynamic program with Bellman equation

$$v(w,e) = \max_{0 \leqslant s \leqslant w} \left\{ r(w,s,e)^\alpha + \beta \left(\sum_{e'} v(s,e')^\gamma \varphi(e') \right)^{\alpha/\gamma} \right\}^{1/\alpha}$$

Here

- w is current wealth (discretized)
- s is savings (discretized)
- e is an IID endowment shock with range E
- β is a constant in $(0,1)$ and r is a reward function

The policy operator corresponding to $\sigma \in \Sigma$ is

$$(T_\sigma\, v)(w,e) = \left\{ r(w,\sigma(w),e)^\alpha + \beta \left(\sum_{e'} v(\sigma(w),e')^\gamma \varphi(e') \right)^{\alpha/\gamma} \right\}^{1/\alpha}$$

Proposition. If

- $\mathsf{X} := \mathsf{W} \times \mathsf{E}$ and
- $V := (0,\infty)^{\mathsf{X}}$,

then $\mathscr{A} = (V, \{T_\sigma\})$ is a max-stable ADP

(Details in Ch 9)

Next consider the operator

$$(B_\sigma\, h)(w) = \left\{ \sum_e \{r(w, \sigma(w), e)^\alpha + \beta h(\sigma(w))^\alpha\}^{\gamma/\alpha}\, \varphi(e) \right\}^{1/\gamma},$$

where h is an element of $(0, \infty)^{\mathsf{W}}$

Define F at $v \in V$ by

$$(Fv)(w) = \left\{ \sum_e v(w, e)^\gamma \varphi(e) \right\}^{1/\gamma} \qquad (w \in \mathsf{W})$$

Then $\mathscr{B} = (F(V), \{B_\sigma\})$ is also an ADP

Moreover, $\mathscr{B}$ is subordinate to $\mathscr{A}$

To see, this, define G_σ by

$$(G_\sigma\, h)(w, e) = \{r(w, \sigma(w), e)^\alpha + \beta h(\sigma(w))^\alpha\}^{1/\alpha}$$

Then

- F and G_σ are order-preserving
- T_σ is equal to $G_\sigma \circ F$ and
- B_σ is equal to $F \circ G_\sigma$

Algorithm 3: Solving $\mathscr{A}$ via $\mathscr{B}$

input $\sigma_0 \in \Sigma$, set $k \leftarrow 0$ and $\varepsilon \leftarrow 1$

while $\varepsilon > 0$ **do**

$h_k \leftarrow$ the fixed point of B_{σ_k}

$\sigma_{k+1} \leftarrow$ an h_k-greedy policy, satisfying

$$\sigma_{k+1}(w) \in \operatorname*{argmax}_{0 \leqslant s \leqslant w} \left\{ \sum_e \left\{ r(w,s,e)^\alpha + \beta h(s)^\alpha \right\}^{\gamma/\alpha} \varphi(e) \right\}^{1/\gamma}$$

$\varepsilon \leftarrow \mathbb{1}\{\sigma_k \neq \sigma_{k+1}\}$ and $k \leftarrow k+1$

end

Compute σ to satisfy

$$\sigma(w,e) \in \operatorname*{argmax}_{0 \leqslant s \leqslant w} \left\{ r(w,s,e)^\alpha + \beta h_k(s)^\alpha \right\}^{1/\alpha}$$

return σ

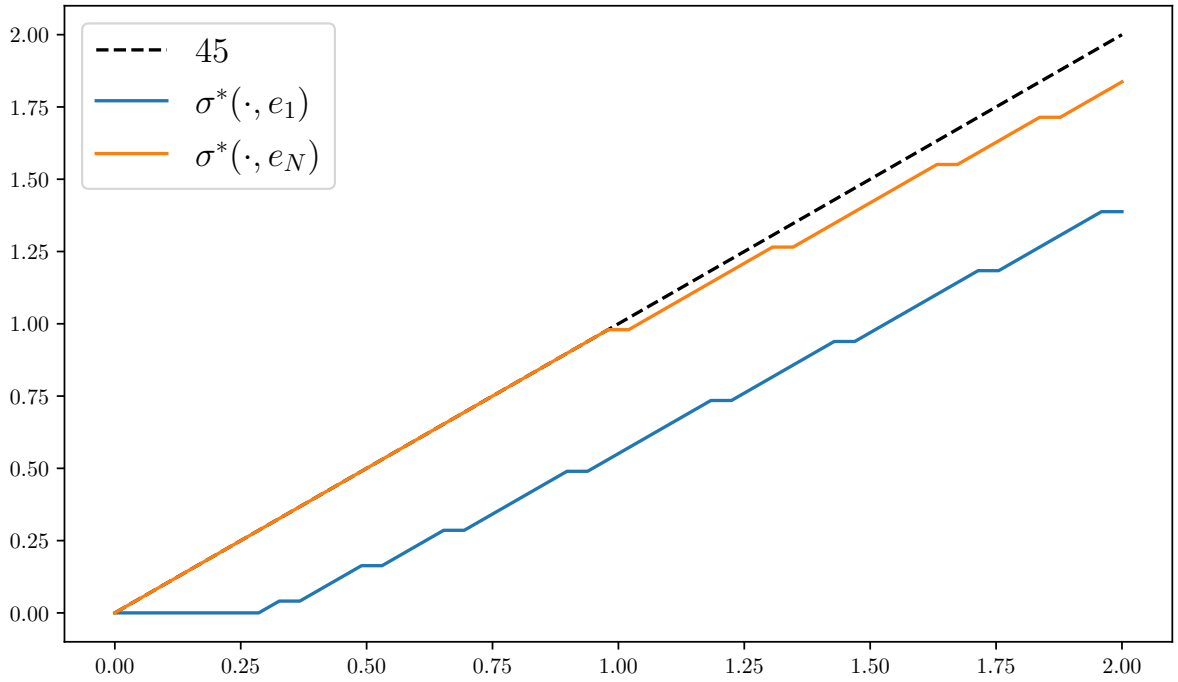


Figure: Optimal savings policy with Epstein–Zin preference

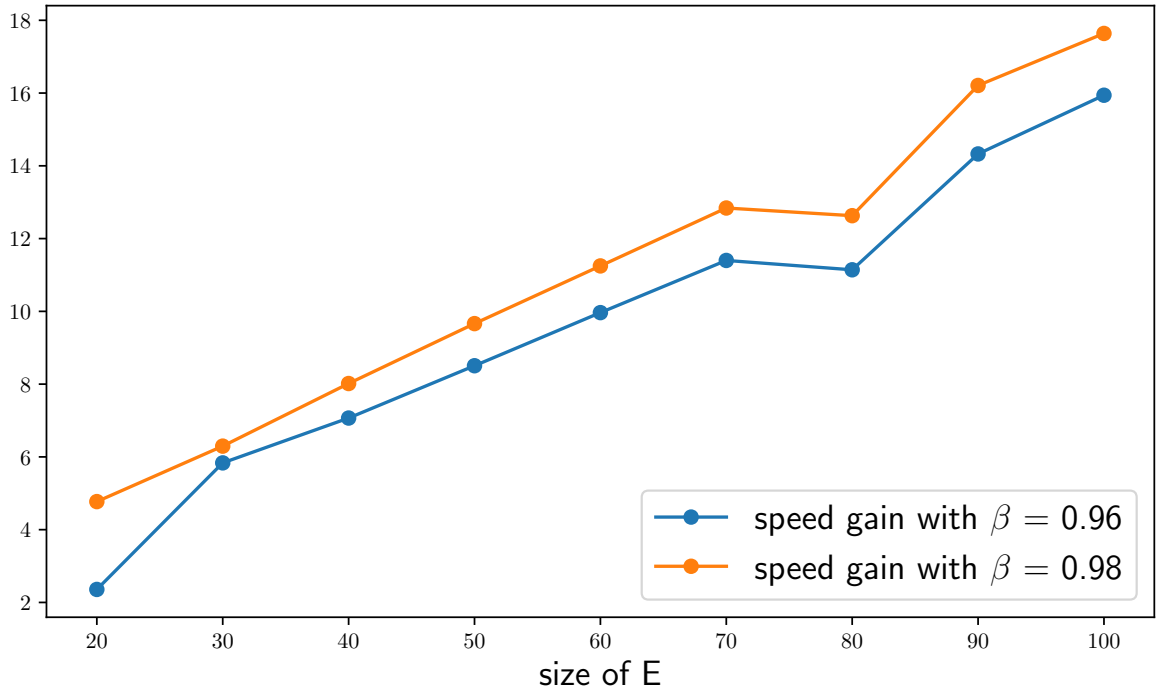


Figure: Speed gain from replacing $\mathscr{A}$ with subordinate model $\mathscr{B}$

For details of computations see

https://github.com/jstac/adps_public