Topics in Computational Economics

Lecture 13

John Stachurski

NYU 2016

Today's Lecture

Dynamic Programming

- Optimality
- Algorithms
- Numerical methods

References

- Stokey and Lucas (1989)
- Stachurski (2009)
- Puterman (1994) Markov Decision Processes

Prequel 1: Bounded Measurable Functions

Let S be a Borel subset of $\mathbb{R}^n$

Let $b\mathbb{R}^S$ be the bounded functions in $\mathbb{R}^S$

Recall that $b\mathbb{R}^S$ is a Banach space with the norm

$$\|f\|_\infty := \sup_{x \in S} |f(x)|$$

Let $b\mathscr{B} :=$ all $\mathscr{B}$-measurable functions in $b\mathbb{R}^S$

This is a **closed** subset of $(b\mathbb{R}^S, \| \cdot \|_\infty)$

Both $b\mathscr{B}$ and $cb\mathbb{R}^S$ are Banach spaces under $\| \cdot \|_\infty$

Prequel 2: General Stochastic Kernels

A **stochastic kernel** on S is a function $P \colon S \times \mathscr{B} \to \mathbb{R}$ such that

1. $x \mapsto P(x, B)$ is $\mathscr{B}$-measurable, for all $B \in \mathscr{B}$
2. $B \mapsto P(x, B)$ is a Borel probability measure, for all $x \in S$

Example. Consider the S-valued process

$$X_{t+1} = F(X_t, \xi_{t+1}) \quad \text{with} \quad \{\xi_t\} \overset{\text{IID}}{\sim} \phi \text{ on } Z$$

The associated stochastic kernel is

$$P(x, B) = \phi\{z \in Z \,:\, F(x, z) \in B\}$$

Each stochastic kernel generates a **conditional expectations operator** $P \colon b\mathscr{B} \to b\mathscr{B}$ defined by

$$Ph(x) = \int h(y) P(x, \mathrm{d}y)$$

Example. The condition expectations operator associated with $X_{t+1} = F(X_t, \xi_{t+1})$ is

$$Ph(x) = \int h[F(x,z)] \phi(dz)$$

The t-th iterate has the interpretation

$$P^t h(x) = \mathbb{E}\left[h(X_t) | X_0 = x\right]$$

Proof for case $t = 2$ is

$$\begin{aligned} P^2 h(x) &= (P(Ph))(x) \\ &= \int (Ph)[F(x,z)]\phi(dz) \\ &= \int \int h[F(F(x,z),z')]\phi(dz')\phi(dz) \\ &= \mathbb{E}\left[h(X_2) | X_0 = x\right] \end{aligned}$$

Fact. P is monotone, in the sense that $f \leqslant g \implies Pf \leqslant Pg$

Fact. P is linear and nonexpansive on $(b\mathscr{B}, \| \cdot \|_\infty)$

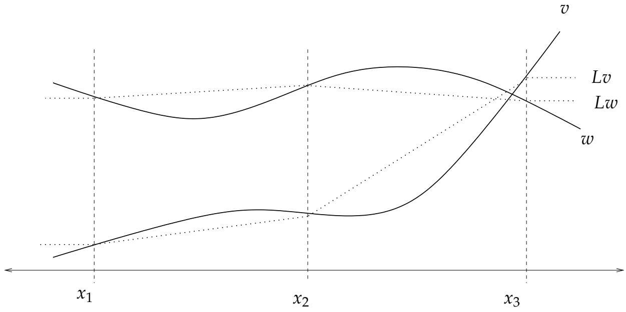
To see that P is nonexpansive, observe that

$$
\begin{aligned}
\|Pf\|_\infty &= \sup_x \left| \int f(y) P(x, \mathrm{d}y) \right| \\
&\leqslant \sup_x \int |f(y)| P(x, \mathrm{d}y) \\
&\leqslant \sup_x \int \|f\|_\infty P(x, \mathrm{d}y) \\
&= \|f\|_\infty \sup_x \int P(x, \mathrm{d}y) = \|f\|_\infty
\end{aligned}
$$

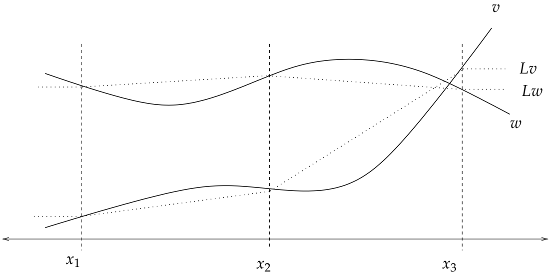
Prequel 3: Nonexpansive Approximation

We can view approximation architectures as operators

Here L maps functions into their piecewise linear approximation



Fact. L is nonexpansive for $\|\cdot\|_\infty$

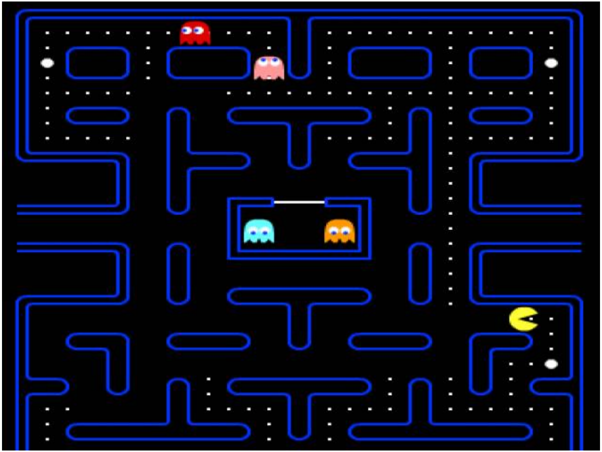


Observe

$$|Lv(x) - Lw(x)| \leqslant \sup_{1 \leqslant i \leqslant k} |v(x_i) - w(x_i)| \leqslant \|v - w\|_\infty$$

Now take supremum over $x \in S$

Markov Decision Processes



Problem: Choose action sequence $\{a_t\}$ to maximize

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t r(X_t, a_t)\right]$$

subject to

- X_{t+1} drawn from $P(X_t, a_t, \mathrm{d}y)$
- $X_0 = x$ given
- $a_t \in \Gamma(X_t)$ for all t

Formally, an **MDP** is a tuple $(S, A, \Gamma, r, \beta, P)$

The components are

- a **state space** S
- an **action space** A
- a **feasible correspondence** $\Gamma \colon S \rightrightarrows A$
- a **reward function** $r \colon \mathbb{G} \to \mathbb{R}$
- a **discount factor** β
- a **stochastic kernel** P from $\mathbb{G}$ to S

Here $\mathbb{G} := \{(x, a) \in S \times A : a \in \Gamma(x)\} =$ graph of Γ

Let's call $\mathbb{G}$ the **feasible state-action pairs**

Interpretation of P:

$$P(x, a, B) = \text{prob } X_{t+1} \in B \text{ when } (x, a) \in \mathbb{G}$$

Fact. Without loss of generality, we can assume that

$$X_{t+1} = F(x, a, \xi_{t+1}) \quad \text{with} \quad \{\xi_t\} \overset{\text{IID}}{\sim} \phi$$

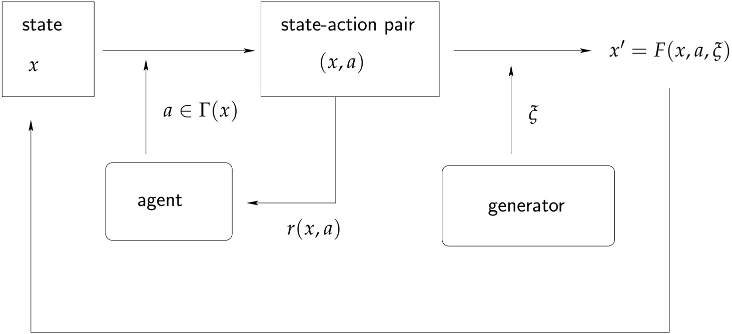
Equivalently,

$$P(x, a, B) = \phi\{z \in Z : F(x, a, z) \in B\}$$

Timing:

1. Agent observes $X_t \in S$
2. Responds with action $a_t \in \Gamma(X_t) \subset A$
3. Receives reward $r(X_t, a_t)$
4. New shock ξ_{t+1} drawn from ϕ
5. X_{t+1} realized as $F(X_t, a_t, \xi_{t+1})$

Now the process repeats


state
x
state-action pair
(x, a)
$x' = F(x, a, \xi)$
$a \in \Gamma(x)$
ξ
agent
$r(x, a)$
generator

Example. Consider the problem

$$\max \mathbb{E}\left[\sum_{t \geqslant 0} \beta^t U(c_t)\right]$$

subject to

$$y_{t+1} = f(y_t - c_t, \xi_{t+1})$$

Here

- the state y_t is a renewable resource
- the action c_t must satisfy $0 \leqslant c_t \leqslant y_t$
- f is a growth function
- $\{\xi_t\}$ is an IID shock sequence

Components

- State space S is $\mathbb{R}_+$
- Action space A is $\mathbb{R}_+$
- Feasible correspondence is $\Gamma(y) = [0, y]$
- $\mathbb{G} = \{(y, c) \in \mathbb{R}^2_+ : 0 \leqslant c \leqslant y\}$
- $r(y, c) = U(c)$
- $P(y, c, B) = \phi\{z \in \mathbb{R}_+ \,:\, f(y - c, z) \in B\}$

Markov Policies

Assume:

- S is a metric space
- A is a metric space
- r and F are Borel measurable
- $\beta \in (0, 1)$
- r is bounded on $\mathbb{G}$

The objective is

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t r(X_t, a_t)\right]$$

To interpret, let's focus on the set of **Markov policies** Σ

Defined as all $\mathscr{B}$-measurable functions $\sigma \colon S \to A$ such that

$$\sigma(x) \in \Gamma(x) \quad \text{for all} \quad x \in S$$

Each $\sigma \in \Sigma$ creates a **controlled Markov process**

$$X_{t+1} = F(X_t, \sigma(X_t), \xi_{t+1})$$

Denoted below as $\{X_t^\sigma\}$ to emphasize dependence on σ

Example. As before, let

$$y_{t+1} = f(y_t - c_t, \xi_{t+1})$$

A **consumption policy** is a map $\sigma \in m\mathscr{B}$ such that

$$0 \leqslant \sigma(y) \leqslant y \qquad (y \in \mathbb{R}_+)$$

Each such policy induces a controlled process

$$y_{t+1} = f(y_t - \sigma(y_t), \xi_{t+1})$$

We write $\{y_t^\sigma\}$ if we need to emphasize dependence on σ

Value of Markov Policies

Define the scalar random variable

$$Y_x^\sigma := \sum_{t \geqslant 0} \beta^t r(X_t^\sigma, \sigma(X_t^\sigma)) \qquad (x = X_0^\sigma)$$

With the notation

$$r_\sigma(x) := r(x, \sigma(x))$$

we have

$$Y_x^\sigma = \sum_{t \geqslant 0} \beta^t r_\sigma(X_t^\sigma)$$

The **policy valuation function** for σ is the function

$$v_\sigma(x) := \mathbb{E}\, Y_x^\sigma \qquad (x \in S)$$

Since r is bounded, by the dominated convergence theorem,

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t r_\sigma(X_t^\sigma)\right] = \sum_{t=0}^{\infty} \beta^t \mathbb{E}\, r_\sigma(X_t^\sigma)$$

That is,

$$v_\sigma(x) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}\, r_\sigma(X_t^\sigma)$$

Operator Theoretic View

Let

$$P_\sigma(x, \mathrm{d}y) := P(x, \sigma(x), \mathrm{d}y)$$

Recall that

$$v_\sigma(x) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}\, r_\sigma(X_t^\sigma)$$

Since $X_0^\sigma = x$, for any h,

$$\mathbb{E}\, h(X_t^\sigma) = P_\sigma^t h(x)$$

Hence

$$v_\sigma = \sum_{t=0}^{\infty} \beta^t P_\sigma^t r_\sigma$$

Fact. v_σ satisfies the functional equation

$$v_\sigma = r_\sigma + \beta P_\sigma v_\sigma$$

Proof:

$$\begin{aligned} v_\sigma &= r_\sigma + \beta P_\sigma r_\sigma + \beta^2 P_\sigma^2 r_\sigma + \cdots \\ &= r_\sigma + \beta P_\sigma \left[r_\sigma + \beta P_\sigma r_\sigma + \cdots \right] = r_\sigma + \beta P_\sigma v_\sigma \end{aligned}$$

Define the **policy valuation operator**

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

By construction, v_σ is a fixed point of T_σ

Computing v_σ

Let $\sigma \in \Sigma$ be given

We know that v_σ is a fixed point of T_σ

If fact

- v_σ is the unique fixed point of T_σ in $b\mathscr{B}$
- $T_\sigma^k w \to v_\sigma$ as $k \to \infty$ for all $w \in b\mathscr{B}$

In particular,

Theorem. T_σ is uniform contraction on $b\mathscr{B}$, with

$$\|T_\sigma w - T_\sigma w'\|_\infty \leqslant \beta \|w - w'\|_\infty \qquad \forall\, w, w' \in b\mathscr{B}$$

Proof: Pick any $x \in S$

Fixing $w, w' \in b\mathscr{B}$, we have

$$
\begin{aligned}
|T_\sigma w(x) - T_\sigma w'(x)| &= \left| \beta \int w(y) P_\sigma(x, \mathrm{d}y) - \beta \int w'(y) P_\sigma(x, \mathrm{d}y) \right| \\
&\leqslant \beta \int |w(y) - w'(y)| P_\sigma(x, \mathrm{d}y) \\
&\leqslant \beta \int \|w - w'\|_\infty P_\sigma(x, \mathrm{d}y) \\
&= \beta \|w - w'\|_\infty \int P_\sigma(x, \mathrm{d}y) \\
&= \beta \|w - w'\|_\infty
\end{aligned}
$$

Now take sup over x

Numerical Methods

To iterate with T_σ in practice we can use an approximation $\hat{T}_\sigma$

Definition of $\hat{T}_\sigma w$:

1. Evaluate $T_\sigma w(x_i)$ for all $x_i \in$ some grid
2. Use a fixed approximation scheme to turn this into $\hat{T}_\sigma w$

Think of step 2 as applying an approximation operator L to $T_\sigma w$

Then $\hat{T}_\sigma$ is the composition $L \circ T_\sigma$

We at iterating with the composition of two operators

Letting $\mathscr{A}$ be the space of approximating functions, we can view it like this

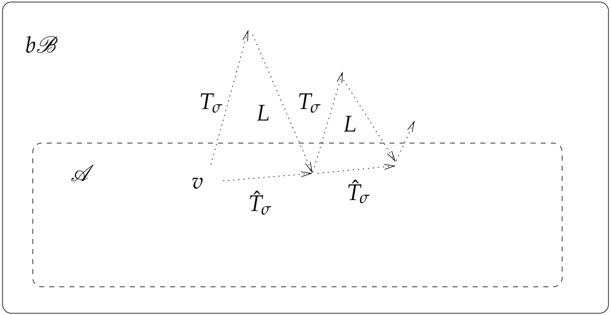


Figure: The map $\hat{T}_\sigma := L \circ T_\sigma$

Fact. If M and N are operators sending metric space (U, d) into itself, N is a uniform contraction with modulus ρ, and M is nonexpansive, then $M \circ N$ is a uniform contraction with modulus ρ

Ex. Prove it

It follows that if L is nonexpansive, then $\hat{T}_\sigma$ is a contraction of modulus β

This gives stability, error bounds, etc.

For details, see, e.g., Stachurski (2009, §10.2.3)

Optimality

Now let's define optimality

First we need some additional assumptions:

1. r is continuous on $\mathbb{G}$
2. $\mathbb{G} \ni (x, a) \mapsto F(x, a, z)$ is continuous for all $z \in Z$
3. $\Gamma(x)$ is continuous and compact-valued for each $x \in S$

Define the **value function** $v^* \colon S \to \mathbb{R}$ by

$$v^*(x) = \sup_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in S) \tag{1}$$

The sup is well defined and finite because

$$|v_\sigma(x)| = \left| \sum_{t=0}^{\infty} \beta^t \mathbb{E}\, r_\sigma(X_t^\sigma) \right| \leqslant \frac{M}{1-\beta} \quad \text{when } M := \sup_{x,a} |r(x,a)|$$

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

$$v_{\sigma^*} = v^*$$

In other words, σ^* attains the sup in (1) for every $x \in S$

Bellman Equation

A function $w \in b\mathscr{B}$ is said to satisfy the **Bellman equation** if

$$w(x) = \sup_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int w(y) P(x,a,\mathrm{d}y) \right\}$$

for all $x \in S$

We might hope that v^* satisfies the Bellman eq, since

- $v^*(y)$ tells us the value of y in terms of discounted rewards
- Varying a in $r(x,a) + \beta \int v^*(y) P(x,a,\mathrm{d}y)$ trades off future vs current rewards
- If we do this optimally we recover $v^*(x)$

We also introduce the **Bellman operator** $T \colon cb\mathbb{R}^S \to cb\mathbb{R}^S$ defined by

$$Tw(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int w(y) P(x,a,\mathrm{d}y) \right\}$$

By construction,

$$w = Tw \iff w \text{ satisfies the Bellman equation}$$

Notes:

- T maps $cb\mathbb{R}^S$ to itself by Berge's theorem of the maximum
- Max exists for all x by Weierstrass's theorem

Greedy Policies

Fix $w \in cb\mathbb{R}^S$

A policy $\sigma \in \Sigma$ is called w-**greedy** if

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int w(y) P(x,a,\mathrm{d}y) \right\}, \quad \forall\, x \in S$$

Equivalent: σ is w-greedy if

$$Tw = T_\sigma w \quad \text{on } S$$

Fact. At least one w-greedy policy exists

Note: Uses a measurable selection theorem to get $\mathscr{B}$-measurability

Theorem. T is a uniform contraction on $(cb\mathbb{R}^S, \|\cdot\|_\infty)$, with

$$\|Tw - Tw'\|_\infty \leqslant \beta \|w - w'\|_\infty, \quad \forall w, w' \in cb\mathbb{R}^S$$

In addition, T is monotone on $cb\mathbb{R}^S$

Monotonicity: The claim is that

$$w, w' \in cb\mathbb{R}^S \text{ and } w \leqslant w' \implies Tw \leqslant Tw'$$

Ex. Check it

Hint: All integrals are monotone

Contraction: Given $w, w' \in cb\mathbb{R}^S$ and $x \in S$, we have

$$
\begin{aligned}
|Tw(x) - Tw'(x)| &= \left| \max_a \left\{ r + \beta \int w \, dP \right\} - \max_a \left\{ r + \beta \int w' \, dP \right\} \right| \\
&\leqslant \beta \max_a \left| \int (w - w') dP \right| \\
&\leqslant \beta \max_a \int |w - w'| dP \\
&\leqslant \beta \max_a \int \|w - w'\|_\infty dP
\end{aligned}
$$

$$
\therefore \quad |Tw(x) - Tw'(x)| \leqslant \beta \|w - w'\|_\infty, \quad \forall x \in S
$$

Now take the sup on the left-hand side

Key Results

Theorem (Blackwell) Under our assumptions, the following statements are true

1. The Bellman equation has exactly one solution in $cb\mathbb{R}^S$
2. That solution is equal to v^*, the value function
3. A policy $\sigma^* \in \Sigma$ is optimal if and only if it is v^*-greedy
4. At least one such policy exists

Remarks:

- 1 is true because T is a contraction
- 2 will be true if $Tv^* = v^*$
- 4 is true by existence of greedy policies

Let w^* be the unique fixed point of T in $cb\mathbb{R}^S$

We claim that $w^* = v^*$

First we show that $w^* \leqslant v^*$

To see this, let $\sigma \in \Sigma$ be w^*-greedy

Then $Tw^* = T_\sigma w^*$

But then $w^* = v_\sigma$, because

- $w^* = Tw^* = T_\sigma w^*$
- v_σ is the only fixed point of T_σ

It follows that $w^* \leqslant v^*$, because $v_\sigma \leqslant v^*$ for all $\sigma \in \Sigma$

Next we show that $v^* \leqslant w^*$

Pick any $\sigma \in \Sigma$

Note that $T_\sigma w^* \leqslant w^*$, because, $\forall \, x \in S$,

$$w^*(x) = Tw^*(x) \geqslant r_\sigma(x) + \beta P_\sigma w^*(x) = T_\sigma w^*(x)$$

Iterating, using monotonicity of T_σ gives

$$T_\sigma^k w^* \leqslant T_\sigma^{k-1} w^* \leqslant \cdots \leqslant T_\sigma^2 w^* \leqslant T_\sigma w^* \leqslant w^*$$

Recall that $T_\sigma^k w^* \to v_\sigma$

Hence taking limits gives $v_\sigma \leqslant w^*$

Since σ is arbitrary it follows that $v^* \leqslant w^*$

Lastly, let's show that σ^* is optimal if and only if it is v^*-greedy

We know that v^* satisfies the Bellman equation, or

$$v^*(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int v^*(y) P(x,a,\mathrm{d}y) \right\}$$

Hence σ^* is v^*-greedy if and only if

$$v^*(x) = r(x, \sigma^*(x)) + \beta \int v^*(y) P(x, \sigma^*(x), \mathrm{d}y)$$

In other words, $v^* = T_{\sigma^*} v^*$

But $v^* \in cb\mathbb{R}^S$, so this is true if and only if $v^* = v_{\sigma^*}$

...which is the definition of optimality

Algorithms

Fitted value function iteration runs as follows:

read in $\{x_i\}_{i=1}^k$, initial $w \in cb\mathbb{R}^S$, and tolerance δ

repeat

evaluate Tw at $\{x_i\}_{i=1}^k$

compute $\hat{T}w = LTw$ from $\{x_i, Tw(x_i)\}_{i=1}^k$

set $e = \|\hat{T}w - w\|_\infty$

set $w = \hat{T}w$

until $e \leqslant \delta$

solve for a w-greedy policy

An alternative is **Howard's policy function iteration** scheme

pick $\sigma \in \Sigma$

repeat

evaluate v_σ

choose $\sigma' \in \Sigma$ such that σ' is v_σ-greedy

set $\sigma = \sigma'$

until a stopping rule is satisfied

Notes

- Make it "fitted" by adding an approximation step
- v_σ can be computed as the fixed point of T_σ

Homework 11

Replicate Fig. 1 of "Stochastic Stability in Monotone Economies"

- https://econtheory.org/ojs/index.php/te/article/view/20140383

Instructions:

- You don't need to produce a 3D graph if you want to show the densities some other way
- Use fitted policy function iteration to solve for optimal policies
- Use the look-ahead estimator to compute stationary densities given the policies
- Submit as a notebook in the usual way