Dynamic Programming

Chapter 4: Optimal Stopping

Thomas J. Sargent and John Stachurski

2024

Topics

- Introduction to optimal stopping
- Theory
- Algorithms
- Applications

Optimal Stopping

Many decision making problems involve choosing when to act

- accept a job
- exit or enter a market
- bring a new product to market
- default on a loan
- exercise a real or financial option

We treat optimal stopping problems via dynamic programming

We begin with the standard theory of optimal stopping

Then we consider alternative approaches

- continuation values
- threshold policies
- etc.

One key objective: provide a rigorous discussion of optimality

This clarifies our intuitive analysis in the context of job search

Theory

Let X be a finite set

An **optimal stopping problem** with state space X consists of

- a stochastic matrix P on X
- a discount factor $\beta \in (0, 1)$
- a **continuation reward function** $c \in \mathbb{R}^{\mathsf{X}}$ and
- an **exit reward function** $e \in \mathbb{R}^{\mathsf{X}}$

Given a P-Markov chain $(X_t)_{t\geqslant 0}$, the problem evolves as follows:

An agent observes X_t each period, continues or stops

If she chooses to stop, she receives $e(X_t)$ and the process ends

If she continues, she receives $c(X_t)$ and the process repeats

Lifetime rewards are given by

$$\mathbb{E} \sum_{t\geqslant 0} \beta^t R_t,$$

where R_t equals

- $c(X_t)$ while the agent continues
- $e(X_t)$ when the agent stops, and zero thereafter

Example. In the infinite-horizon job search problem

- wage offer process (W_t) is $\stackrel{\text{IID}}{\sim} \varphi$ on finite W
- stop $=$ accept offer
- continue $=$ receive unemployment compensation, repeat

This is an optimal stopping problem with

- state space $\mathsf{X} = \mathsf{W}$
- stochastic matrix P with all rows equal to φ
- exit reward function $e(x) = x/(1 - \beta)$ and
- the continuation reward function $\equiv$ unemployment compensation

Example. Infinite-horizon American call option

Provides the right to buy a given asset at strike price K at every future date

The market price of the asset is given by $S_t = s(X_t)$

- (X_t) is P-Markov on finite set X and $s \in \mathbb{R}^{\mathsf{X}}$
- the interest rate is $r > 0$

When to exercise is an optimal stopping problem, with

- discount factor $\beta = 1/(1+r)$
- exit reward function $e(x) = s(x) - K$ and
- continuation reward zero

A **policy function** is a map σ from X to $\{0, 1\}$

Interpretation: observe state x, respond with action $\sigma(x)$, where

- $\sigma(x) = 0$ means "continue"
- $\sigma(x) = 1$ means "stop"

State must contain enough information to decide!

Let $\Sigma =$ all $\sigma \colon \mathsf{X} \to \{0, 1\}$

Let $v_\sigma(x) =$ expected lifetime value of following policy σ now and forever, given current state $x \in \mathsf{X}$

We call v_σ the **σ-value function**

Our aim:

- choose a policy that maximizes lifetime value!

In particular, $\sigma^* \in \Sigma$ is called **optimal** if

$$v_{\sigma^*}(x) = \max_{\sigma \in \Sigma} v_\sigma(x) \quad \text{for all } x \in \mathsf{X}$$

Before we can compute optimal policies, we need to be able to evaluate v_σ for each $\sigma \in \Sigma$

- How can we do this?

Some thought will convince you that v_σ must satisfy

$$v_\sigma(x) = \sigma(x)e(x) + (1 - \sigma(x)) \left[c(x) + \beta \sum_{x' \in \mathsf{X}} v_\sigma(x') P(x, x') \right]$$

Case 1: $\sigma(x) = 1$

Then the above states $v_\sigma(x) = e(x)$, which is the right value

Case 2: $\sigma(x) = 0$

Then

$$v_\sigma(x) = c(x) + \beta \sum_{x' \in \mathsf{X}} v_\sigma(x') P(x, x')$$

A natural recursion

To repeat, we need to solve

$$v_\sigma(x) = \sigma(x)e(x) + (1 - \sigma(x)) \left[c(x) + \beta \sum_{x' \in \mathsf{X}} v_\sigma(x') P(x, x') \right]$$

Let

- $r_\sigma(x) := \sigma(x)e(x) + (1 - \sigma(x))c(x)$
- $L_\sigma(x, x') := \beta(1 - \sigma(x))P(x, x')$

Then the equation becomes $v_\sigma = r_\sigma + L_\sigma \, v_\sigma$

Ex. Show that $r(L_\sigma) < 1$ and hence

$$v_\sigma = (I - L_\sigma)^{-1} \, r_\sigma$$

We can also view v_σ as the fixed point of the **policy operator**

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

Pointwise this is

$$T_\sigma v = r_\sigma + L_\sigma v$$

We know that v_σ is the unique solution to $v = r_\sigma + L_\sigma v$

Hence v_σ is the unique fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$

Ex. Prove that T_σ is order-preserving on $\mathbb{R}^{\mathsf{X}}$

Ex. Show that T_σ is a contraction map on $\mathbb{R}^{\mathsf{X}}$

Given $f, g \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$, we have

$$|(T_\sigma f)(x) - (T_\sigma\, g)(x)| = \left| (1 - \sigma(x))\beta \sum_{x'} (g(x') - f(x'))P(x, x') \right|$$

$$\leqslant \beta \left| \sum_{x'} [f(x') - g(x')] P(x, x') \right|$$

Applying the triangle inequality and $\sum_{x' \in \mathsf{X}} P(x, x') = 1$ yields

$$|(T_\sigma f)(x) - (T_\sigma\, g)(x)| \leqslant \beta \|f - g\|_\infty$$

Hence

$$\|T_\sigma f - T_\sigma\, g\|_\infty \leqslant \beta \|f - g\|_\infty$$

Ex. Show that T_σ is a contraction map on $\mathbb{R}^{\mathsf{X}}$

Given $f, g \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$, we have

$$|(T_\sigma f)(x) - (T_\sigma\, g)(x)| = \left| (1 - \sigma(x))\beta \sum_{x'} (g(x') - f(x'))P(x, x') \right|$$

$$\leqslant \beta \left| \sum_{x'} [f(x') - g(x')] P(x, x') \right|$$

Applying the triangle inequality and $\sum_{x' \in \mathsf{X}} P(x, x') = 1$ yields

$$|(T_\sigma f)(x) - (T_\sigma\, g)(x)| \leqslant \beta \|f - g\|_\infty$$

Hence

$$\|T_\sigma f - T_\sigma\, g\|_\infty \leqslant \beta \|f - g\|_\infty$$

We define the **value function** of the optimal stopping problem as

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

Thus, $v^*(x) =$ max lifetime value given current state x

Next steps

1. introduce **the Bellman equation**

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

2. prove that the Bellman equation has a unique solution in $\mathbb{R}^{\mathsf{X}}$
3. show that this solution equals v^*

The **Bellman operator** for the optimal stopping is defined by

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

When we discuss optimality the next Ex. will be useful

Ex. Prove that T is an order preserving self-map on $\mathbb{R}^{\mathsf{X}}$

Proof: Fix $f, g \in \mathbb{R}^{\mathsf{X}}$ with $f \leqslant g$

Since $P \geqslant 0$, we have $Pf \leqslant Pg$

Hence $c + \beta P f \leqslant c + \beta P g$

$$\therefore \quad Tf = e \vee (c + \beta P f) \leqslant e \vee (c + \beta P g) = Tg$$

The **Bellman operator** for the optimal stopping is defined by

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

When we discuss optimality the next Ex. will be useful

Ex. Prove that T is an order preserving self-map on $\mathbb{R}^{\mathsf{X}}$

Proof: Fix $f, g \in \mathbb{R}^{\mathsf{X}}$ with $f \leqslant g$

Since $P \geqslant 0$, we have $Pf \leqslant Pg$

Hence $c + \beta Pf \leqslant c + \beta Pg$

$$\therefore \quad Tf = e \vee (c + \beta Pf) \leqslant e \vee (c + \beta Pg) = Tg$$

Ex. Prove that, for all $\sigma \in \Sigma$, T dominates T_σ on $\mathbb{R}^{\mathsf{X}}$

Proof: For all $a, b \in \mathbb{R}$ and $\lambda \in [0, 1]$, we have

$$\lambda a + (1 - \lambda) b \leqslant a \vee b$$

Now fix $\sigma \in \Sigma$, $v \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$

We have

$$(T_\sigma v)(x) = \sigma(x) e(x) + (1 - \sigma(x)) \left[c(x) + \beta \sum_{x'} v(x') P(x, x') \right]$$

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

Ex. Prove that, for all $\sigma \in \Sigma$, T dominates T_σ on $\mathbb{R}^{\mathsf{X}}$

Proof: For all $a, b \in \mathbb{R}$ and $\lambda \in [0,1]$, we have

$$\lambda a + (1 - \lambda) b \leqslant a \vee b$$

Now fix $\sigma \in \Sigma$, $v \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$

We have

$$(T_\sigma v)(x) = \sigma(x) e(x) + (1 - \sigma(x)) \left[c(x) + \beta \sum_{x'} v(x') P(x, x') \right]$$

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

Value function vs Bellman equation

Prop. Under the stated conditions

1. T is a contraction map of modulus β on $\mathbb{R}^{\mathsf{X}}$ and
2. the unique fixed point of T on $\mathbb{R}^{\mathsf{X}}$ is the value function v^*

In particular, v^* is the unique solution to the Bellman equation

The first step of the proof is the next exercise

Ex. Given f, g in $\mathbb{R}^{\mathsf{X}}$, show that

$$\|Tf - Tg\|_\infty \leqslant \beta \|f - g\|_\infty$$

Proof: Recall the bound $|z \vee a - z \vee b| \leqslant |a - b|$

From this we have

$$
\begin{aligned}
|Tf - Tg| &= |e \vee (c + \beta P f) - [e \vee (c + \beta P g)]| \\
&\leqslant |c + \beta P f - (c + \beta P g)| \\
&= \beta \, |P(f - g)| \\
&\leqslant \beta P \, |f - g|
\end{aligned}
$$

$$
\therefore \quad |(Tf)(x) - (Tg)(x)| \leqslant \beta \sum_{x'} |f(x') - g(x')| P(x, x')
$$

$$
\therefore \quad \|Tf - Tg\|_\infty \leqslant \beta \|f - g\|_\infty
$$

Now we know T has a unique fixed point $\bar{v}$ in $\mathbb{R}^{\mathsf{X}}$

Next we claim that

$$\bar{v} = v^*$$

We show

- $\bar{v} \leqslant v^*$ and
- $\bar{v} \geqslant v^*$

To prove $\bar{v} \leqslant v^*$, define $\sigma \in \Sigma$ by

$$\sigma(x) := \mathbb{1}\left\{e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} \bar{v}(x')P(x, x')\right\}$$

For this choice of σ, for any $x \in \mathsf{X}$,

$$(T_\sigma \bar{v})(x) = \sigma(x)e(x) + (1 - \sigma(x))\left[c(x) + \beta \sum_{x' \in \mathsf{X}} \bar{v}(x')P(x, x')\right]$$

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

$$= (T\bar{v})(x) = \bar{v}(x)$$

In particular,

$$T_\sigma \, \bar v = \bar v$$

But the only fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$ is the σ-value function v_σ!

Hence $\bar v = v_\sigma$

But then $\bar v \leqslant v^*$, by the definition of v^*

- why?

It only remains to prove $\bar v \geqslant v^*$

Fix $\sigma \in \Sigma$

Since

1. T dominates T_σ on $\mathbb{R}^{\mathsf{X}}$ and
2. T is order-preserving and globally stable

we have $v_\sigma \leqslant \bar{v}$

- why?

Taking the max over $\sigma \in \Sigma$ implies $v^* \leqslant \bar{v}$

We have now proved that v^* is the unique solution to the Bellman equation!

Finding optimal policies

For $v \in \mathbb{R}^{\mathsf{X}}$, we call $\sigma \in \Sigma$ ***v*-greedy** if

$$\forall\, x \in \mathsf{X}, \quad \sigma(x) = \mathbb{1}\left\{ e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, x') \right\}$$

- treats v as the value function and optimizes

Ex. Show that $\sigma \in \Sigma$ is v^*-greedy if and only if $T_\sigma v^* = v^*$

Proof: We have

$$\sigma \in \Sigma \text{ is } v^*\text{-greedy} \iff \sigma e + (1 - \sigma)(c + \beta P v^*) = e \vee (c + \beta P v^*)$$

$$\iff T_\sigma v^* = v^*$$

Finding optimal policies

For $v \in \mathbb{R}^{\mathsf{X}}$, we call $\sigma \in \Sigma$ v-**greedy** if

$$\forall\, x \in \mathsf{X}, \quad \sigma(x) = \mathbb{1}\left\{ e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, x') \right\}$$

- treats v as the value function and optimizes

Ex. Show that $\sigma \in \Sigma$ is v^*-greedy if and only if $T_\sigma v^* = v^*$

Proof: We have

$$\sigma \in \Sigma \text{ is } v^*\text{-greedy} \iff \sigma e + (1 - \sigma)(c + \beta P v^*) = e \vee (c + \beta P v^*)$$

$$\iff T_\sigma v^* = v^*$$

Prop. $\sigma \in \Sigma$ is optimal $\iff$ σ is v^*-greedy

Proof: For $\sigma \in \Sigma$, the following are all equivalent

1. σ is v^*-greedy
2. $T_\sigma v^* = v^*$
3. $v^* = v_\sigma$

- Why are 2 and 3 equivalent?

This result is a version of **Bellman's principle of optimality**

Prop. $\sigma \in \Sigma$ is optimal $\iff$ σ is v^*-greedy

Proof: For $\sigma \in \Sigma$, the following are all equivalent

1. σ is v^*-greedy
2. $T_\sigma v^* = v^*$
3. $v^* = v_\sigma$

- Why are 2 and 3 equivalent?

This result is a version of **Bellman's principle of optimality**

Corollary. The optimal stopping problem has exactly one optimal policy

Proof: For each $v \in \mathbb{R}^{\mathsf{X}}$, the greedy policy

$$\sigma^*(x) := \mathbb{1}\left\{ e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, x') \right\} \qquad (x \in \mathsf{X})$$

is uniquely defined

By the last Proposition, a policy is optimal iff it is v^*-greedy

Hence exactly one optimal policy exists

Corollary. The optimal stopping problem has exactly one optimal policy

Proof: For each $v \in \mathbb{R}^{\mathsf{X}}$, the greedy policy

$$\sigma^*(x) := \mathbb{1}\left\{ e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, x') \right\} \qquad (x \in \mathsf{X})$$

is uniquely defined

By the last Proposition, a policy is optimal iff it is v^*-greedy

Hence exactly one optimal policy exists

Firm Valuation with Exit

Previously we discussed firm valuation under various scenarios

Value was defined as expected present value of profit stream

- a standard and popular methodology
- easy to apply

But it ignores an important fact

- firms have the option to exit and sell remaining assets

Now we consider firm valuation in the presence of this exit option

We consider a firm where

- $\pi_t = \pi(Z_t)$ for some fixed $\pi \in \mathbb{R}^{\mathsf{Z}}$
- $(Z_t)_{t \geqslant 0}$ is Q-Markov on finite set $\mathsf{Z} \subset \mathbb{R}$

At the start of each period, the firm decides whether to

- remain in operation, receiving current profit π_t, or
- exit, receiving $s > 0$ for sale of all assets

Discounting is at fixed rate r and $\beta := 1/(1 + r)$

We assume that $r > 0$

Let Σ be all $\sigma \colon \mathsf{Z} \to \{0, 1\}$

For given $\sigma \in \Sigma$ and $v \in \mathbb{R}^{\mathsf{Z}}$, the policy operator is

$$(T_\sigma v)(z) = \sigma(z)s + (1 - \sigma(z) \left[\pi(z) + \beta \sum_{z'} v(z')Q(z, z') \right]$$

Recall that

- T_σ has a unique fixed point v_σ
- $v_\sigma(z) :=$ the value of following policy σ forever, given $Z_0 = z$

The Bellman operator is the order-preserving self-map T on $\mathbb{R}^{\mathsf{Z}}$ defined by

$$(Tv)(z) = \max \left\{ s, \pi(z) + \beta \sum_{z'} v(z') Q(z, z') \right\}$$

Pointwise, $Tv = s \vee (\pi + \beta Q v)$

Let v^* be the value function for this problem

- v^* is the unique fixed point of T in $\mathbb{R}^{\mathsf{Z}}$
- successive approximation from any $v \in \mathbb{R}^{\mathsf{Z}}$ converges to v^*
- a policy is optimal if and only if it is v^*-greedy