Quantitative Economics Workshop Paris

Dynamic Programming

John Stachurski

September 2022

Introduction

Summary of this lecture:

- Introduce dynamic programming
- Introduce the RDP framework and provide examples
  - inventories, optimal savings, job search
- Provide RDP optimality results
- Discuss algorithms
- Show algorithms and their performance for some case
  - optimal savings, optimal investment...

Dynamic program

an initial state X_0 is given
$t \leftarrow 0$
while $t < T$ **do**
  observe current state X_t
  choose action A_t
  receive reward R_t based on (X_t, A_t)
  state updates to X_{t+1}
  $t \leftarrow t + 1$
end

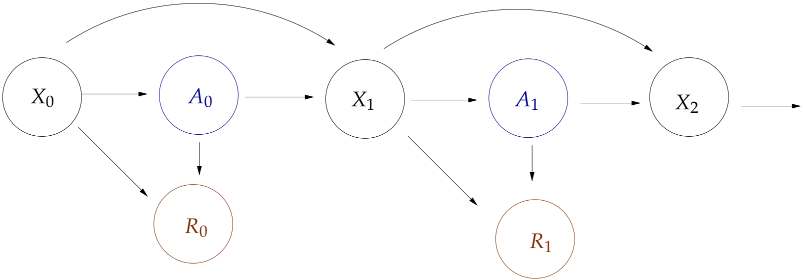

Figure: A dynamic program

Comments:

- Objective: maximize **lifetime rewards**
  - Example. $\mathbb{E}[R_0 + \beta R_1 + \beta^2 R_2 + \cdots]$ for some $\beta \in (0, 1)$
- If $T < \infty$ then the problem is called a **finite horizon** problem
- Otherwise it is called an **infinite horizon** problem
- The update rule can also depend on random elements:

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

Example: Optimal Inventories

Given a demand process $(D_t)_{t \geqslant 0}$, inventory $(X_t)_{t \geqslant 0}$ obeys

$$X_{t+1} = F(X_t, A_t, D_{t+1})$$

where

- A_t is units of stock ordered this period
- $F(x, a, d) := \max\{x - d, 0\} + a$

The firm can store at most K items at one time

- The state space is $\mathsf{X} := \{0, \ldots, K\}$

We assume $(D_t) \stackrel{\text{IID}}{\sim} \varphi \in \mathcal{D}(\mathbb{Z}+)$

Profits are given by

$$\pi_t := X_t \wedge D_{t+1} - cA_t - \kappa \mathbb{1}\{A_t > 0\}$$

- Orders in excess of inventory are lost
- c is unit product cost (and unit sales prices $= 1$)
- κ is a fixed cost of ordering inventory

With $\beta := 1/(1+r)$ and $r > 0$, the value of the firm is

$$V_0 = \mathbb{E} \sum_{t \geqslant 0} \beta^t \pi_t$$

Objective: maximize (shareholder) value

Expected current profit is

$$r(x,a) := \sum_{d \geqslant 0} (x \wedge d)\varphi(d) - ca - \kappa \mathbb{1}\{a > 0\}$$

The set of feasible order sizes at x is

$$\Gamma(x) := \{0, \ldots, K - x\}$$

- Γ is the **feasible correspondence**

The **Bellman equation** is

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{d \geqslant 0} F(x,a,d)\varphi(d) \right\}$$

The **standard solution procedure** for this problem is VFI:

1. define the Bellman operator T via

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_d v(F(x,a,d)) \varphi(d) \right\}$$

2. iterate with T to (approximately) compute the fixed point v^* and

3. compute a v^*-greedy greedy policy σ^*, which satisfies

$$\sigma^*(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_d v^*(F(x,a,d)) \varphi(d) \right\}$$

Optimal Savings with Labor Income

Wealth evolves according to

$$C_t + W_{t+1} \leqslant R W_t + Y_t \qquad (t = 0, 1, \ldots)$$

- (W_t) takes values in finite set $\mathsf{W} \subset \mathbb{R}_+$
- (Y_t) is Q-Markov chain on finite set Y
- $C_t \geqslant 0$

The household maximizes

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

The Bellman equation is

$$v(w,y) = \max_{w' \in \Gamma(w,y)} \left\{ u(Rw + y - w') + \beta \sum_{y' \in \mathsf{Y}} v(w',y') Q(y,y') \right\}$$

The standard solution procedure is VFI

1. Set up Bellman operator T
2. Iterate with T from some initial guess to approximate v^*
3. Compute the v^*-greedy policy

Recursive Decision Processes

We will study an abstract dynamic program with Bellman equation

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

Advantages of "abstract" dynamic programming

- Subsumes standard Markov decision processes
- Can handle state-dependent discounting, recursive prefs, etc.
- Abstraction means clean proofs
- Abstraction allows better analysis of algorithms

Let X and A be finite sets (**state** and **action spaces**)

Actions are constrained by a nonempty correspondence Γ from X to A called the **feasible correspondence**

The feasible correspondence in turn defines

1. the **feasible state-action pairs**

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

2. the set of **feasible policies**

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

- "follow" $\sigma \iff$ always respond to state x with action $\sigma(x)$

Given X, A and Γ, a **recursive decision process** (RDP) consists of

1. a closed subset $\mathcal{V}$ of $\mathbb{R}^{\mathsf{X}}$ called the set of **candidate value functions** and
2. a **value aggregator**, which is a function

$$B \colon \mathsf{G} \times \mathcal{V} \to \mathbb{R}$$

satisfying $v, w \in \mathcal{V}$ and $v \leqslant w \implies$

$$B(x, a, v) \leqslant B(x, a, w) \text{ for all } (x, a) \in \mathsf{G}$$

and

$$\sigma \in \Sigma \text{ and } v \in \mathcal{V} \implies w \in \mathcal{V} \quad \text{where } w(x) := B(x, \sigma(x), v)$$

Example. For the inventory problem we set

- $\Gamma(x) := \{0, \ldots, K - x\}$,
- $\mathcal{V} = \mathbb{R}^{\mathsf{X}}$ and

$$B(x, a, v) := \left\{ r(x, a) + \beta \sum_{d \geqslant 0} v(F(x, a, d)) \varphi(d) \right\}$$

The Bellman equation is then

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

The function B is a valid value aggergator

For example, if $v \leqslant w$, then

$$B(x, a, v) \leqslant B(x, a, w)$$

Example. For the savings problem we set

- $\Gamma(w, y) := \{w' \in \mathsf{W} : w' \leqslant Rw + y\}$,
- $\mathcal{V} = \mathbb{R}^{\mathsf{X}}$ and

$$B((w, y), w', v) := u(Rw + y - w') + \beta \sum_{y' \in \mathsf{Y}} v(w', y') Q(y, y')$$

The Bellman equation is then

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

The function B is a valid value aggergator

For example, if $f \leqslant g$, then

$$B((w, y), w', f) \leqslant B((w, y), w', g)$$

Discuss possible generalizations

- state-dependent discounting
- recursive preferences

Operators

Given v in $\mathcal{V}$, we call $\sigma \in \Sigma$ **v-greedy** if

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} B(x, a, v) \qquad \text{for all } x \in \mathsf{X}$$

The **Bellman operator** is defined by

$$(Tv)(x) = \max_{a \in \Gamma(x)} B(x, a, v) \qquad (x \in \mathsf{X},\ v \in \mathcal{V})$$

- v^* solves the Bellman equation iff v^* is a fixed point of T

For each $\sigma \in \Sigma$, the **policy operator** T_σ is

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

Example. The policy operator for the savings problem, given $\sigma \in \Sigma$, is

$$(T_\sigma v)(w, y) =$$

$$u(Rw + y - \sigma(w, y)) + \beta \sum_{y' \in \mathsf{Y}} v(\sigma(w, y), y') Q(y, y')$$

Stability

Let $\mathcal{R} := (\Gamma, \mathcal{V}, B)$ be an RDP with

- Bellman operator T and
- policy operators $\{T_\sigma\}_{\sigma \in \Sigma}$

We call $\mathcal{R}$ **globally stable** if

1. T is globally stable on $\mathcal{V}$ and
2. T_σ is globally stable on $\mathcal{V}$ for all $\sigma \in \Sigma$

Example. In the inventory problem,

Lifetime Value

For a globally stable RDP, given $\sigma \in \Sigma$, let v_σ be the unique solution to

$$v_\sigma(x) = B(x, \sigma(x), v_\sigma) \quad \text{for all } x \in \mathsf{X}$$

- the unique fixed point of T_σ

Key idea: $v_\sigma =$ lifetime value of following the policy σ in each period

Example. In the inventory problem, where

$$v_\sigma(x) = r(x, \sigma(x)) + \beta \sum_d v_\sigma(F(x, \sigma(x), d)\varphi(d)$$