

# INTRO TO DYNAMIC PROGRAMMING

---

QUANTITATIVE ECONOMICS 2024

Piotr Żoch

December 5, 2024

## A TYPICAL PROBLEM

- Many problems seen in economics have a common structure:
  - We observe the current **state**  $X_t$  at time  $t$ .
  - We choose an **action**  $A_t$  at time  $t$ .
  - We get a **reward**  $R_t$  at time  $t$ .
  - The state progresses to  $X_{t+1}$  at time  $t + 1$ .
- If the largest possible  $t$  is  $T < \infty$ , then we have a **finite horizon** problem. Otherwise we have an **infinite horizon** problem.

## EXAMPLE

- You saw a simple lifecycle model with known income path (*perfect foresight*).
- It was easy to solve: just calculate optimal consumption and savings in every period.  
Possible to do it by hand.
- What if income earned by households is uncertain?
- What if households can invest in assets with uncertain returns?
- We need to find optimal solution for all possible paths of income and returns.

## EXAMPLE

- Consider a problem of a firm that produces a good. The firm wants to maximize the expected present discounted value of profits:

$$\mathbb{E} \sum_{t=0}^{\infty} \left( \frac{1}{1+r} \right) \pi_t$$

- $X_t$  is the current state of the firm. It can be the current level of capital, the current level of inventory, prices set by competitors...
- $A_t$  is the action taken by the firm. It can be the level of production, the level of future inventory, the price of the good...
- $R_t$  is the reward. Here it is the profit of the firm,  $\pi_t$ .
- $X_{t+1}$  is the state of the firm in the next period. It can depend on the current state  $X_t$  and the action  $A_t$  taken.

## EXAMPLE

- This is potentially an **extremely** complicated problem.
- For example: the state  $X_t$  can include the demand for the good – and it could be random.
- Find actions for all possible future states...
- We will learn tools that can help us solve such problems.

## PLAN

- Today we will study an example: **McCall's job search model (1970)**.
- Exposition based on Stachurski and Sargent (2023).
- Next time: more general theory of dynamic programming.

## MCCALL'S JOB SEARCH MODEL

## TWO-PERIOD PROBLEM

- An unemployed agent receives a job offer at wage  $W_t$ .
- She can either **accept** the offer or **reject** it.
- If she **accepts**, she gets this wage permanently.
- If she **rejects**, she gets unemployment benefit  $c$ .
- Wage offers are independent and identically distributed (i.i.d.) and nonnegative, with distribution  $\phi$ :
  - $W \subset \mathbb{R}_+$  is a finite set of possible wages.
  - $\phi : W \rightarrow [0, 1]$  is a probability mass function,  $\phi(w)$  is the probability of getting a wage  $w$ .
- The agent is risk-neutral and impatient: the utility of getting  $y$  tomorrow is  $\beta y$ , with  $\beta \in (0, 1)$ .



## TWO-PERIOD PROBLEM

- The agent lives for two periods ( $t = \{1, 2\}$ ) and starts unemployed.
- The question: is it better to accept a received offer or wait until tomorrow hoping for better offer?
- What is the lowest wage that the agent should accept?
- We will start analyzing the problem by looking at the second period,  $t = 2$ : **backward induction**.

## PERIOD $T=2$

- Suppose the agent is unemployed at  $t = 2$ .
- She gets a wage offer  $W_2$  which she can either **accept** or **reject** the offer.
- **Accept**: get income  $W_2 \rightarrow$  get utility  $W_2$ .
- **Reject**: get income  $c \rightarrow$  get utility  $c$ .
- Since this is the last period of her life, she will accept if and only if

$$W_2 \geq c.$$

## PERIOD T=1

- The agent gets a wage offer  $W_1$ .
- She can either (a) accept and get  $W_1$  forever, or (b) reject and get  $c$  in period  $t = 1$  and then get  $\max \{W_2, c\}$  in period  $t = 2$ .
- The utility of (a) is  $W_1 + \beta W_1$ . We call it the stopping value.
- The utility of (b) is  $h_1 := c + \beta \mathbb{E} \max \{W_2, c\}$ . We call it the continuation value.

$$h_1 = c + \beta \sum_{w' \in W} v_2(w') \phi(w'), \quad v_2(w') := \max \{w', c\}$$

- The agent will accept if and only if the stopping value is greater than the continuation value:

$$W_1 + \beta W_1 \geq h_1.$$

## VALUE FUNCTION

- The key object in dynamic programming is the **value function**.
- It is a **function** that maps the state to the **maximum** expected present discounted value of future rewards.
- In our example, there are two stages: time  $t$  and the received wage offer  $w$ .
- $v_2(w)$  is the value function at time  $t = 2$  and wage  $w$ : the largest possible reward that the agent can get if she starts unemployed at  $t = 2$  and gets a wage offer  $w$ . We have

$$v_2(w) = \max \{w, c\}.$$

- The time 1 value function is

$$v_1(w) := \max \left\{ w + \beta w, c + \beta \sum_{w' \in W} v_2(w') \phi(w') \right\}.$$

## TWO-PERIOD EXAMPLE

- This particular problem is easy to solve.
- Accept if

$$w \geq \frac{h_1}{1 + \beta}$$

so the value function is

$$v_1(w) = \begin{cases} (1 + \beta) w & \text{if } w \geq \frac{h_1}{1 + \beta} \\ h_1 & \text{otherwise.} \end{cases}$$

- In context of this example, we call  $w^* := \frac{h_1}{1 + \beta}$  the **reservation wage**.
- We see that since  $h_1$  is increasing in  $c$ , the reservation wage is higher when the unemployment benefit  $c$  is higher.

## THREE-PERIOD EXAMPLE

- Extend the model by one period,  $t = 0$ .
- The value function at  $t = 0$  is

$$v_0(w) := \max \left\{ w + \beta w + \beta^2 w, c + \beta \sum_{w' \in W} v_1(w') \phi(w') \right\}.$$

where the formula for  $v_1$  is from the previous slide.

- **Key insight:** at  $t = 0$  it is like a two-period problem.
- **All** information about the future is summarized in  $v_1$ , the value function at  $t = 1$ .
- This is the standard approach: convert a complicated dynamic optimization problem into a sequence of two-period problems.

## BELLMAN EQUATION

- Recall we had

$$v_2(w) = \max \{w, c\}$$

$$v_1(w) = \max \left\{ w + \beta w, c + \beta \sum_{w' \in W} v_2(w') \phi(w') \right\}$$

$$v_0(w) = \max \left\{ w + \beta w + \beta^2 w, c + \beta \sum_{w' \in W} v_1(w') \phi(w') \right\}.$$

- The recursive relationships between the value functions are called the **Bellman equations**.
- Warning:** these equations are (in general) **functional equations**. We need to find functions, not numbers. Here it is easy: we have a finite set of possible wages – treat functions as vectors.

## INFINITE HORIZON

- The three-period problem is also easy to solve.
- In fact, we can use the same approach (backward induction) as before for any finite horizon problem.
- What if the horizon is **infinite**? We no longer have the **terminal** period.
- Dynamic programming makes this problem **tractable**.



## INFINITE HORIZON

- The objective function is

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

where  $R_t \in \{c, W_t\}$ .

- Let  $\beta \in (0, 1)$  be the discount factor, as before we let  $c > 0$ .
- The wage process satisfies  $(W_t)_{t \geq 0} \stackrel{iid}{\sim} \phi$  where  $\phi \in \mathcal{D}(W)$  and  $W \subset \mathbb{R}_+$  with  $|W| < \infty$ .
- For any finite or countable set  $F$ ,  $\mathcal{D}(F)$  is the set of distributions on  $F$ .

## INFINITE HORIZON

- What is **stopping value**?
- If the worker accepts wage  $w$  she gets

$$w + \beta w + \beta^2 w + \beta^3 w + \dots = \frac{w}{1 - \beta}.$$

- What is the **continuation value**?
- If the worker rejects wage  $w$  she gets

$$c + \beta \sum_{w' \in W} v(w') \phi(w').$$

note that the value function is the same in all periods – there is always infinite remaining future.

## INFINITE HORIZON

- Bellman equation is:

$$v(w) = \max \left\{ \frac{w}{1 - \beta}, c + \beta \sum_{w' \in W} v(w') \phi(w') \right\}.$$

- **Principle of optimality, Bellman (1960):** *An optimal policy has the property that whatever the initial state and the initial decisions it must constitute an optimal policy with regards to the state resulting from the first decision.*
- This is not that trivial, we will return to it (and prove it!) later.
- Interpretation: value function satisfies the Bellman equation.

## CHALLENGE

- Bellman equation is:

$$v(w) = \max \left\{ \frac{w}{1 - \beta}, c + \beta \sum_{w' \in W} v(w') \phi(w') \right\}.$$

- Once we have  $v(w)$  we can characterize the optimal choice of the agent.
- Q: how to find  $v(w)$ ? It is a function!
- Q: is there a solution?
- Q: is the solution unique?
- Q: what are the properties of the solution?
- A: we will learn how to answer these questions.

## APPROACH I

- This particular problem is relatively easy.
- Recall how we defined the **continuation value**:

$$h^* := c + \beta \sum_{w' \in W} \max \left\{ \frac{w'}{1 - \beta}, h^* \right\} \phi(w')$$

- We can write the value function as

$$v(w) = \max \left\{ \frac{w}{1 - \beta}, h^* \right\}.$$

- **Key:**  $h^*$  is a scalar, not a function. When would it break down?
- Find  $h^*$  directly, by solving the equation numerically.
- Then use  $h^*$  to get the value function.

## APPROACH I

---

### Algorithm Solving for $v$ directly

---

```
1: procedure MCCALL
2:    $k \leftarrow 1, \epsilon \leftarrow \tau + 1, h_k \leftarrow c$ 
3:   while  $\epsilon > \tau$  do
4:      $h_{k+1} \leftarrow c + \beta \sum_{w' \in W} \max \left\{ \frac{w'}{1-\beta}, h_k \right\} \phi(w')$ 
5:      $\epsilon \leftarrow |h_{k+1} - h_k|, k \leftarrow k + 1$ 
6:   end while
7:   for  $w \in W$  do
8:      $v(w) \leftarrow \max \left\{ \frac{w}{1-\beta}, h_k \right\}$ 
9:   end for
10: end procedure
```

---

## APPROACH II

- We know how to solve a finite horizon problem – backward induction.
- Maybe we can get an approximate solution to the infinite horizon problem by considering a finite horizon problem with a very large number of periods?
- We will **prove** that it actually works.

## MATHEMATICAL DETOUR: BANACH'S CONTRACTION MAPPING THEOREM



## EXISTENCE AND UNIQUENESS

- Before solving any problem it is useful to know if there is a solution at all.
- We will use a powerful theorem that will help us answer this question.
- First we introduce a concept of a **fixed point**.
- Let  $U$  be any nonempty set. We call  $T$  a self-map on  $U$  if  $T : U \rightarrow U$ .
- For a self-map  $T$  on  $U$ , we say that a point  $u^* \in U$  is a **fixed point** of  $T$  if  $Tu^* = u^*$ .

## FIXED POINT

- Some examples of fixed points:
  - $U = \mathbb{R}$ ,  $T(u) = 2u + 3$ . Then  $u^* = -3$  is a fixed point of  $T$ .
  - $U = [0, 1]$ ,  $T(u) = u$ . Then every  $u \in U$  is a fixed point of  $T$ .
  - $U = \mathbb{R}$ ,  $T(u) = u + 1$ . There is no fixed point of  $T$ .
- **Global stability:** a self-map  $T$  on  $U$  is **globally stable** on  $U$  if  $T$  has a unique fixed point  $u^*$  in  $U$  and  $T^k u \rightarrow u^*$  for all  $u \in U$ .

## METRIC SPACE

- A metric space is a set  $U$ , together with a metric (distance function)  $\rho, \rho : U \times U \rightarrow \mathbb{R}$ , such that for all  $u, v, w \in U$ :
  - $\rho(u, v) \geq 0$  (nonnegativity), with equality if and only if  $u = v$ .
  - $\rho(u, v) = \rho(v, u)$  (symmetry).
  - $\rho(u, v) \leq \rho(u, w) + \rho(w, v)$  (triangle inequality).

## METRIC SPACE

- **Convergence**: a sequence  $\{x_n\}_{n=0}^{\infty}$  in  $U$  **converges** to  $x \in U$ , if for each  $\epsilon > 0$ , there exists  $N_{\epsilon}$  such that  $\rho(x_n, x) < \epsilon, \forall n \geq N_{\epsilon}$ .
- **Cauchy sequence**: A sequence  $\{x_n\}_{n=0}^{\infty}$  in  $U$  is **Cauchy** if for each  $\epsilon > 0$ , there exists  $N_{\epsilon}$  such that  $\rho(x_n, x_m) < \epsilon, \forall n, m \geq N_{\epsilon}$ .
- **Complete metric space**: A metric space  $(U, \rho)$  is **complete** if every **Cauchy sequence** in  $U$  **converges** to an element in  $U$ .
- Example:  $\mathbb{R}$  with  $\rho(u, v) = |u - v|$  is a complete metric space.

## NORMED VECTOR SPACE

- A normed vector space is a vector space  $V$  over  $\mathbb{R}$  or  $\mathbb{C}$  together with a norm  $\|\cdot\|, \|\cdot\| : V \rightarrow \mathbb{R}$ , such that for all  $u, v \in V$  and  $\alpha \in \mathbb{R}$  or  $\mathbb{C}$ :
  - $\|u\| \geq 0$  (nonnegativity), with equality if and only if  $u = 0$ .
  - $\|\alpha u\| = |\alpha| \|u\|$  (absolute homogeneity).
  - $\|u + v\| \leq \|u\| + \|v\|$  (triangle inequality).
- Some norms:  $\|u\|_1 = \sum_{i=1}^n |u_i|$  (Manhattan),  $\|u\|_2 = \left(\sum_{i=1}^n u_i^2\right)^{1/2}$  (Euclidean),  $\|u\|_\infty = \max_{i=1, \dots, n} |u_i|$  (supremum).
- We will only focus on real vector spaces.
- A normed vector space is a metric space with  $\rho(u, v) = \|u - v\|$ .
- A complete normed vector space is called a **Banach space**.

## NORMED VECTOR SPACE

- Let  $X \subseteq \mathbb{R}^n$  be a nonempty set, let  $C(X)$  be the set of **bounded continuous functions** on  $X$  with the supremum norm,  $\|f\|_{\infty} = \sup_{x \in X} |f(x)|$ . Then  $C$  is a Banach space (complete normed vector space).

## CONTRACTION

- Let  $(U, \rho)$  be a **metric space** and  $T$  a self-map on  $U$ .  $T$  is a **contraction mapping** (with modulus  $\lambda$ ) if for some  $\lambda \in (0, 1)$

$$\rho(Tu, Tv) \leq \lambda \rho(u, v), \quad \text{for all } u, v \in U.$$

- If  $T$  is a contraction on  $U$ , then  $T$  is uniformly continuous on  $U$ .

## BANACH'S CONTRACTION MAPPING THEOREM

### **Theorem (Banach's contraction mapping theorem)**

*If  $(U, \rho)$  is a complete metric space and a self-map  $T$  is a contraction mapping with modulus  $\lambda$ , then:*

- $T$  has a unique fixed point  $u^*$  in  $U$ , and*
- for any  $u_0 \in U$ ,  $\rho(T^k u_0, u^*) \leq \lambda^k \rho(u_0, u^*)$  for all  $k \in \mathbb{N}$ .*



## BANACH'S CONTRACTION MAPPING THEOREM

- By the contraction property of  $T$  we have:

$$\rho(u_2, u_1) = \rho(Tu_1, Tu_0) \leq \lambda \rho(u_1, u_0).$$

By induction:

$$\rho(u_{k+1}, u_k) \leq \lambda^k \rho(u_1, u_0), n = 1, 2, \dots$$

Using it and the triangle inequality, for  $m > n$

$$\begin{aligned} \rho(u_m, u_n) &\leq \rho(u_m, u_{m-1}) + \rho(u_{m-1}, u_{m-2}) + \dots + \rho(u_{n+1}, u_n) \\ &\leq \left[ \lambda^{m-1} + \lambda^{m-2} + \dots + \lambda^n \right] \rho(u_1, u_0) \\ &\leq \frac{\lambda^n}{1 - \lambda} \rho(u_1, u_0). \end{aligned}$$

## PROOF

- From

$$\rho(u_{k+1}, u_k) \leq \frac{\lambda^k}{1 - \lambda} \rho(u_1, u_0)$$

we see that  $\{u_k\}_{k=0}^{\infty}$  is a **Cauchy sequence**.

- Since  $U$  is complete,  $\{u_k\}_{k=0}^{\infty}$  converges to some  $u^* \in U$ .
- We now show that  $u^*$  is a **fixed point** of  $T$ . For all  $n$  and all  $u_0 \in U$ ,

$$\begin{aligned} \rho(Tu^*, u^*) &\leq \rho(Tu^*, T^n u_0) + \rho(T^n u_0, u^*) \\ &\leq \lambda \rho(u^*, T^{n-1} u_0) + \rho(T^n u_0, u^*). \end{aligned}$$

- By the previous result both terms on the right hand side go to 0 as  $n \rightarrow \infty$ .

## PROOF

- We need to show there is no other  $\hat{u}$  such that  $T\hat{u} = \hat{u}$ .
- Suppose there is such  $\hat{u} \neq u^*$ . Take  $\rho(\hat{u}, u^*) = \delta > 0$  Then

$$\delta = \rho(u^*, \hat{u}) = \rho(Tu^*, T\hat{u}) \leq \lambda \rho(u^*, \hat{u}) = \lambda \delta.$$

but  $\delta \leq \delta \lambda$  **cannot** hold because  $\lambda < 1$ !

## PROOF

- To prove "for any  $u_0 \in U$ ,  $\rho(T^k u_0, u^*) \leq \lambda^k \rho(u_0, u^*)$  for all  $k \in \mathbb{N}$ " notice that for any  $n \geq 1$ :

$$\rho(T^k u_0, u^*) = \rho(T(T^{k-1} u_0), T u^*) \leq \lambda \rho(T^{k-1} u_0, u^*).$$

- We can also show that

$$\rho(T^k u_0, u^*) \leq \frac{1}{1-\lambda} \rho(T^k u_0, T^{k-1} u_0).$$

## BANACH'S CONTRACTION MAPPING THEOREM

- Banach's contraction mapping theorem is a very powerful result.
- First, we can use it to **show** that a **unique solution exists** (fixed point).
- Second, it gives us a way to **find** the fixed point – we can iterate the contraction mapping (successive approximation / fixed point iteration)
- It proves that the fixed point iteration converges. It also gives us a bound on the rate of convergence ( $\lambda$ , the modulus of contraction).
- We often use it to prove the existence of a solution to a **functional equation** or to find a **stationary distribution**.

## AN EXAMPLE

- Let  $U = \mathbb{R}$  and  $\|\cdot\| = |\cdot|$ , and  $T$  be a self-map on  $U$  defined by  $T(u) = 0.5u + 3$ .
- $U$  with  $\rho(u, v) = |u - v|$  is a Banach space.
- We have

$$\begin{aligned}\rho(T(u), T(v)) &= |T(u) - T(v)| \\ &= |0.5u + 3 - 0.5v - 3| \\ &= 0.5 \cdot |u - v| \\ &= 0.5 \cdot \rho(u, v).\end{aligned}$$

so  $T$  is a contraction with modulus  $\lambda = 0.5$ .

- By the CMT, there exists a unique fixed point of  $T$  on  $U$ :  $u^* = 6$ .

## AN EXAMPLE

- Let  $U = \mathbb{R}$  and  $\|\cdot\| = |\cdot|$ , and  $T$  be a self-map on  $U$  defined by  $T(u) = 0.5u + 3$ .
- $U$  with  $\rho(u, v) = |u - v|$  is a Banach space.
- We have

$$\begin{aligned}\rho(T(u), T(v)) &= |T(u) - T(v)| \\ &= |0.5u + 3 - 0.5v - 3| \\ &= 0.5 \cdot |u - v| \\ &= 0.5 \cdot \rho(u, v).\end{aligned}$$

so  $T$  is a contraction with modulus  $\lambda = 0.5$ .

- By the CMT, there exists a unique fixed point of  $T$  on  $U$ :  $u^* = 6$ .

## AN EXAMPLE

- Another application: Picard-Lindelöf theorem.
- Suppose we have an initial value problem:

$$y'(t) = f(t, y(t)), \quad y(t_0) = y_0.$$

- If  $f(t, \cdot)$  is continuous and bounded and  $f(t, \cdot)$  is Lipschitz continuous in  $y$  with Lipschitz constant  $L$  for every  $t \in [t_0 - \alpha, t_0 + \alpha]$ , then there exists a unique solution to the problem in the neighborhood of  $t_0$ .



BACK TO JOB SEARCH MODEL

## BELLMAN EQUATION

- We want to solve the Bellman equation:

$$v(w) = \max \left\{ \frac{w}{1 - \beta}, c + \beta \sum_{w' \in W} v(w') \phi(w') \right\}.$$

- Introduce a **Bellman operator** defined at  $v \in \mathbb{R}^W$  as

$$(Tv)(w) := \max \left\{ \frac{w}{1 - \beta}, c + \beta \sum_{w' \in W} v(w') \phi(w') \right\}.$$

- Note: here we treat  $v$  as a vector, not a function. We can do it, because  $W$  is finite.
- Let  $V := \mathbb{R}_+^W$  and let  $\|\cdot\|_\infty$  be the supremum norm. Note that  $V$  with this norm is a Banach space.

## BELLMAN EQUATION

- Notice that

$$|\max \{a, x\} - \max \{a, y\}| \leq |x - y| \quad \text{for all } a, x, y \in \mathbb{R}.$$

- Take any  $f, g \in V$  and fix any  $w \in W$ . Use the above to get

$$\begin{aligned} |(Tf)(w) - (Tg)(w)| &\leq \beta \left| \sum_{w' \in W} [f(w') - g(w')] \phi(w') \right| \\ &\leq \beta \|f - g\|_{\infty} \end{aligned}$$

- Take the supremum over  $w$  to get

$$\|Tf - Tg\|_{\infty} \leq \beta \|f - g\|_{\infty}.$$

- This proves  $T$  is a contraction with modulus  $\beta$ .

## BELLMAN EQUATION

- There exists a unique fixed point  $v^*$  of  $T$  in  $V$ .
- The fixed point of the Bellman operator solves the Bellman equation.
- The solution to the Bellman equation is a fixed point of the Bellman operator.
- We can obtain  $v^*$  by iterating the Bellman operator:

$$v_{k+1} = Tv_k, \quad k = 0, 1, \dots$$

- We can start with **any**  $v_0 \in V$ .

## APPROACH II

---

### Algorithm Value function iteration

---

```
1: procedure VFI
2:    $k \leftarrow 1, \epsilon \leftarrow \tau + 1, v_k \leftarrow v_{init}$ 
3:   while  $\epsilon > \tau$  do
4:     for  $w \in W$  do
5:        $v_{k+1}(w) \leftarrow (Tv_k)(w)$ 
6:     end for
7:      $\epsilon \leftarrow \|v_{k+1} - v_k\|_\infty, k \leftarrow k + 1$ 
8:   end while
9: end procedure
```

---

## OPTIMAL CHOICES

- Once we have  $v^*$  we can characterize the optimal choice of the agent.
- We can calculate the continuation value  $h^*$ :

$$h^* := c + \beta \sum_{w' \in W} v^*(w') \phi(w').$$

- Reject the offer if  $w / (1 - \beta) < h^*$ , accept otherwise.
- Denote rejection given wage  $W_t$  as  $A_t = 0$  and acceptance as  $A_t = 1$ .

## OPTIMAL CHOICES

- Let  $A_t = \sigma_t(W_t)$  be the optimal choice of the agent at time  $t$  given wage offer  $W_t$ .
- We call  $\sigma_t$  the (time  $t$ ) **policy function**.
- In this particular case the policy function is

$$\sigma_t(w) = \begin{cases} 1 & \text{if } \frac{w}{1-\beta} \geq h^* \\ 0 & \text{otherwise.} \end{cases}$$

- The policy function here depends only on the current state (wage).
- We call such a policy function (depending on the current state only) **Markov policy**.

## OPTIMAL CHOICES

- A policy is an "instruction manual" for the agent: what to do in each state.
- For an agent following  $\sigma \in \Sigma$ , if the current wage offer is  $w$ , the agent will respond with  $\sigma(w) \in \{0, 1\}$ .
- For each  $v \in V$ , a  **$v$  – greedy policy** is a  $\sigma \in \Sigma$  satisfying

$$\sigma(w) = \mathbb{1} \left\{ \frac{w}{1 - \beta} \geq c + \beta \sum_{w' \in W} v(w') \phi(w') \right\} \quad \text{for all } w \in W.$$

- The recommendation is: **adopt a  $v^*$  – greedy policy** (notice the superscript!).
- This is a restatement of **Bellman's principle of optimality**.



## LOOKING FORWARD

- Here we have a **finite** state space – we can treat  $v$  as a vector.
- We only used a fraction of the power of dynamic programming...
- What if we move **away** from finite state and action spaces?
- What are the conditions under which the Bellman operator is a **contraction**? Is there an easy way to check it?
- Does the **principle of optimality** hold in **general**?