Dynamic Programming: From Local Optimality to Global Optimality

John Stachurski, Jingni Yang and Humphrey Yang

July 9, 2025

Introduction

DP notation:

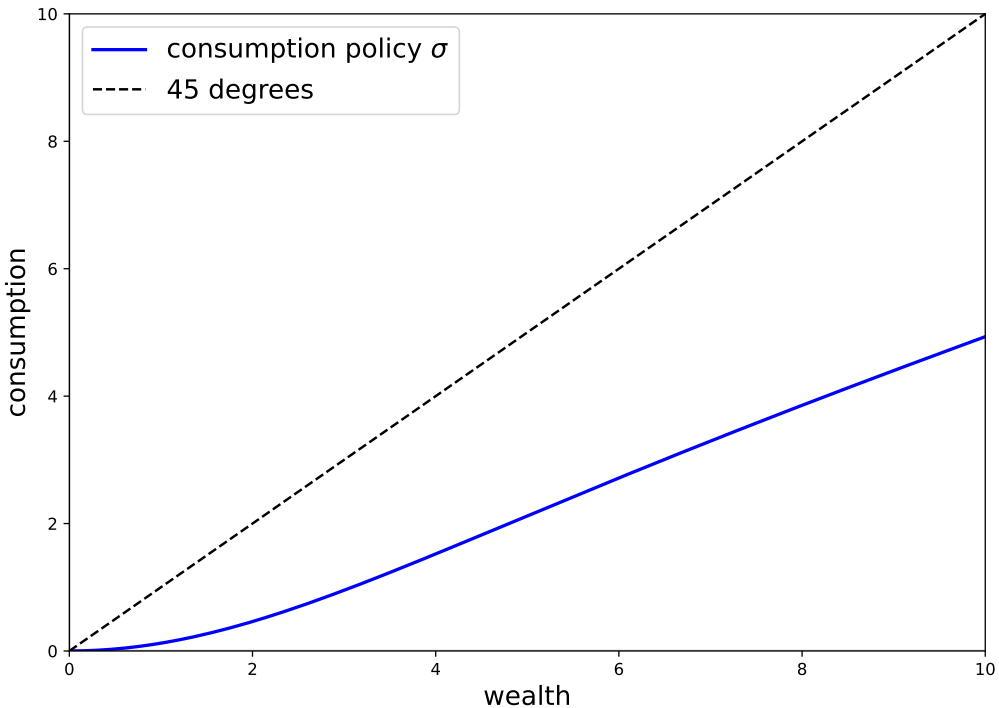
- X is the **state space**
- A is the **action space**
- A **policy** is a map σ: X $\to$ A
- $\sigma(x)$ = action in state x

A **feasible policy** is a policy satisfying some feasibility constraint

We let Σ denote the **set of all feasible policies**

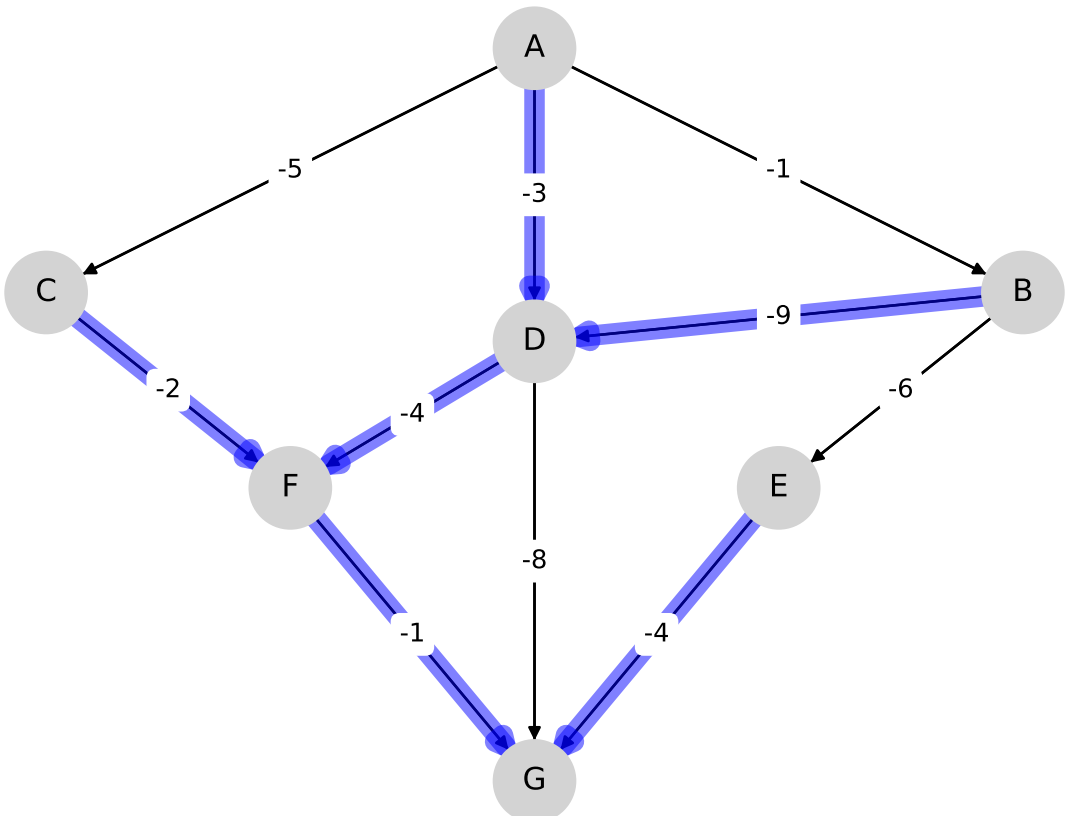
Eg. Optimal savings

- $X = A = \mathbb{R}_+$
- σ maps wealth to consumption



Eg. Traversing a graph

- X = all nodes and A = all edges
- σ maps each node to an edge



Eg. User engagement

A controller feeds media content to a given user

- Objective = maximize user attention
- Action = next video in feed
- State = is user's history, measures of engagement

$|A| = 10^{?}$

$|X| = 10^{????}$

Let $v_\sigma(x)$ = **lifetime value** when

1. following policy σ forever
2. starting from initial condition $x \in \mathsf{X}$

Eg. In the optimal savings problem

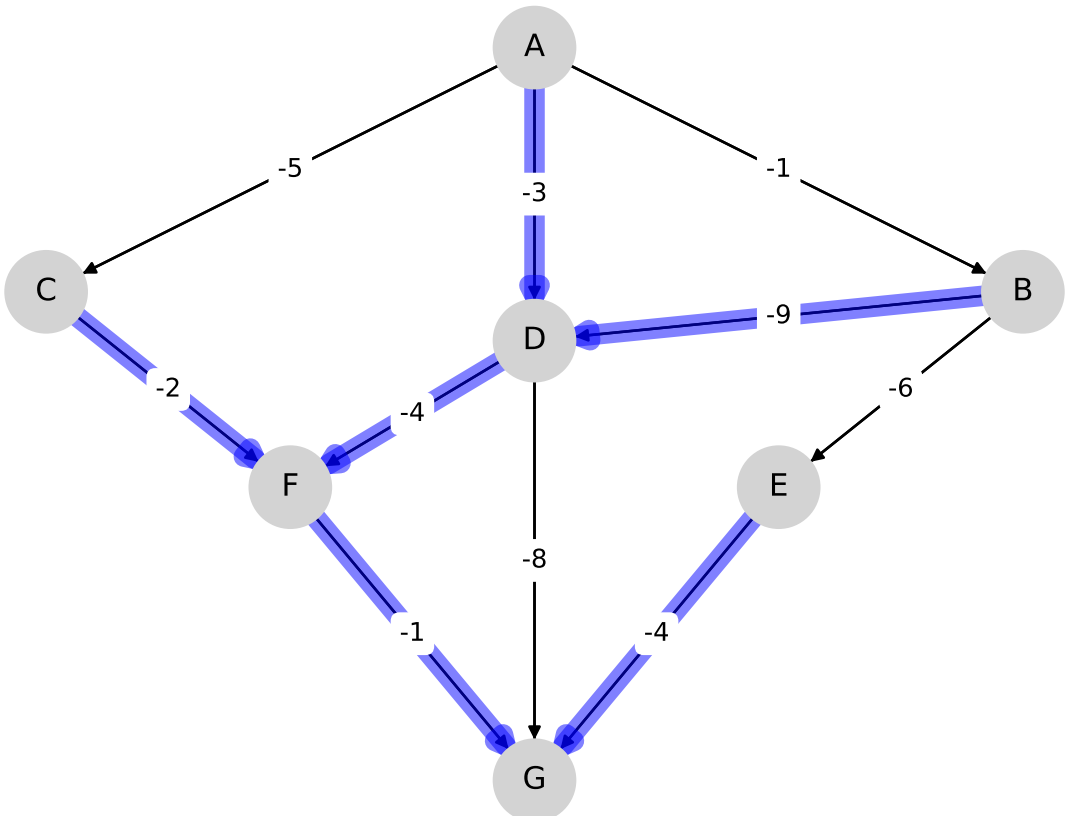
$$v_\sigma(x) = \mathbb{E} \sum_{t \geqslant 0} \beta^t u(C_t)$$

where

- $C_t = \sigma(W_t)$
- $W_0 = x$
- $W_{t+1} = R_{t+1}(W_t - \sigma(W_t)) + Y_{t+1}$ for $t = 0, 1, \ldots$

Eg. $v_\sigma(x)$ = reward from traversing graph starting at node x

- $v_\sigma(A) = -8$



The DP problem: find and characterize optimal policies

Def. Policy σ is called **optimal** when

$$v_{\sigma}(x) = \max_{s \in \Sigma} v_s(x) \text{ for } \textbf{every } x \in \mathsf{X}$$

Equivalent:

$$v_\sigma = v^* \qquad \text{where} \qquad v^*(x) := \sup_{\sigma \in \Sigma} v_\sigma(x)$$

Equivalent:

$$v_s \leqslant v_\sigma \text{ for all } s \in \Sigma$$

The DP problem: find and characterize optimal policies

Def. Policy σ is called **optimal** when

$$v_\sigma(x) = \max_{s \in \Sigma} v_s(x) \text{ for } \textbf{every } x \in \mathsf{X}$$

Equivalent:

$$v_\sigma = v^* \qquad \text{where} \qquad v^*(x) := \sup_{\sigma \in \Sigma} v_\sigma(x)$$

Equivalent:

$$v_s \leqslant v_\sigma \text{ for all } s \in \Sigma$$

The DP problem: find and characterize optimal policies

Def. Policy σ is called **optimal** when

$$v_\sigma(x) = \max_{s \in \Sigma} v_s(x) \text{ for } \textbf{every } x \in \mathsf{X}$$

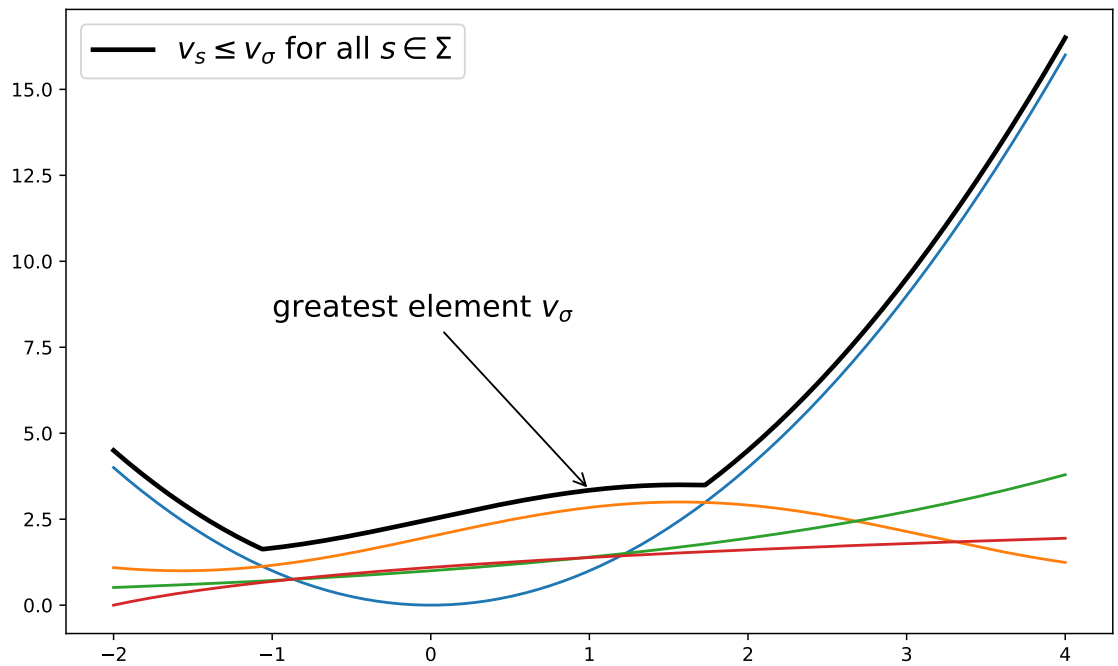
Equivalent:

$$v_\sigma = v^* \qquad \text{where} \qquad v^*(x) := \sup_{\sigma \in \Sigma} v_\sigma(x)$$

Equivalent:

$$v_s \leqslant v_\sigma \text{ for all } s \in \Sigma$$

In other words, we seek a **greatest element** in a **partially ordered set**



Our research question

Recall: σ is optimal when

$$v_{\sigma}(x) = v^*(x) \text{ for } \textbf{every } x \in \mathsf{X}$$

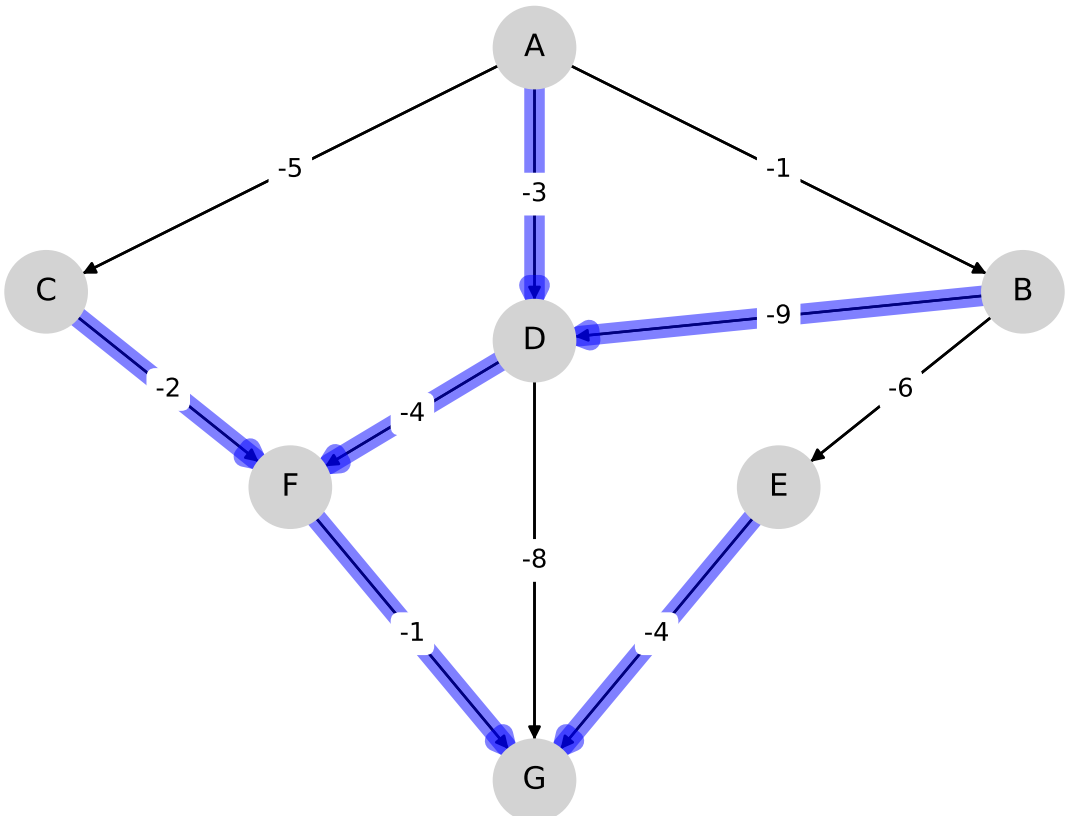
Our question: Under what conditions do we have

$$v_{\sigma}(x) = v^*(x) \text{ for } \textbf{some } x \in \mathsf{X} \implies v_{\sigma}(x) = v^*(x) \text{ for } \textbf{every } x \in \mathsf{X}$$

When does "local optimality" imply "global optimality"?

Not always true, of course

- $v_\sigma(A) = v^*(A)$ but not $v_\sigma(B) = v^*(B)$



When might it be true?

- Say we freeze x and solve $\max_\sigma v_\sigma(x)$
- The value $v_\sigma(x)$ depends on present and future rewards
- Hence making $v_\sigma(x)$ large requires that σ does well in the future
- Therefore, **σ should perform well in states we are likely to visit**
- If we visit all of X under σ, then σ should perform well everywhere

Possible conclusion:

Sufficient mixing under $\sigma \implies$

local optimality implies global optimality

Motivation

1. Better understanding of theoretical properties
2. Assist modern solution methods for high-dimensional DP problems

Motivation

1. Better understanding of theoretical properties
2. Assist modern solution methods for high-dimensional DP problems

Example: Policy gradient methods

1. Fix an **initial distribution** ρ and set

$$F(\sigma) \coloneqq \int v_\sigma(x)\rho(\mathrm{d}x)$$

2. Maximize F over Σ using **gradient ascent**:

$$\sigma_{n+1} = \sigma_n + \lambda_n \nabla F(\sigma_n)$$

Key point: Fixing an initial distribution $\implies$ **real-valued objective**

$$F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$$

Example: Policy gradient methods

1. Fix an **initial distribution** ρ and set

$$F(\sigma) \coloneqq \int v_\sigma(x)\rho(\mathrm{d}x)$$

2. Maximize F over Σ using **gradient ascent**:

$$\sigma_{n+1} = \sigma_n + \lambda_n \nabla F(\sigma_n)$$

Key point: Fixing an initial distribution $\implies$ **real-valued objective**

$$F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$$

In practice,

- Replace Σ with $\{\sigma(\cdot, \theta) : \theta \in \Theta\}$ where $\sigma(\cdot, \theta)$ is an ANN
- Replace the objective function with

$$F(\theta) := \int v_{\sigma(\cdot,\theta)}(x)\rho(\mathrm{d}x)$$

- Now replace F with a Monte Carlo aproximation $\hat{F}$

Notice that we are just training an ANN with loss function $-\hat{F}$

Benefit: Engineering teams have extremely powerful tools for training ANNs in high dimensions

- autodiff, backprop, JIT compilers, 1000s of GPUs, etc.

In practice,

- Replace Σ with $\{\sigma(\cdot, \theta) : \theta \in \Theta\}$ where $\sigma(\cdot, \theta)$ is an ANN
- Replace the objective function with

$$F(\theta) := \int v_{\sigma(\cdot,\theta)}(x)\rho(\mathrm{d}x)$$

- Now replace F with a Monte Carlo aproximation $\hat{F}$

Notice that we are just training an ANN with loss function $-\hat{F}$

Benefit: Engineering teams have extremely powerful tools for training ANNs in high dimensions

- autodiff, backprop, JIT compilers, 1000s of GPUs, etc.

But does maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ lead to optimality?

Special case: does maximizing $F(\sigma) = v_\sigma(\bar{x})$ lead to optimality?

In other words, **does local optimality imply global optimality?**

- Are there settings where this holds true?

But does maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ lead to optimality?

Special case: does maximizing $F(\sigma) = v_\sigma(\bar{x})$ lead to optimality?

In other words, **does local optimality imply global optimality?**

- Are there settings where this holds true?

But does maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ lead to optimality?

Special case: does maximizing $F(\sigma) = v_\sigma(\bar{x})$ lead to optimality?

In other words, **does local optimality imply global optimality?**

- Are there settings where this holds true?

Time for theory...

Setup

Consider a DP problem with **Bellman equation**

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

Here

- $\beta \in (0, 1)$ is the **discount factor**
- Γ is the **feasible correspondence**
- r is the **reward function**
- P is the **transition kernel**

As before

- The **value function** is defined by $v^*(x) := \sup_{\sigma \in \Sigma} v_\sigma(x)$
- A policy σ is called **optimal** if $v_\sigma = v^*$

For given σ we let

$$P_\sigma(x, \mathrm{d}x') = P(x, \sigma(x), \mathrm{d}x') = \textbf{Markov kernel under } \sigma$$

We define the **Bellman operator** by

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

Given σ we introduce the **policy operator**

$$(T_\sigma\, v)(x) := r(x,\sigma(x)) + \beta \int v(x') P(x,\sigma(x),\mathrm{d}x')$$

Facts:

- The value function v^* is the unique fixed point of T
- The lifetime value v_σ is the unique fixed point of T_σ

Standard optimality results hold

Proposition 1

Under mild assumptions,

1. At least one optimal policy exists
2. v^* is the unique solution to the Bellman equation
3. A policy σ is optimal if and only if

$$\sigma(x) \in \arg\max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \int v^*(x') P(x, a, \mathrm{d}x') \right\} \quad \text{for all } x \in \mathsf{X}$$

4. Value function iteration converges to v^*

Our results

Consider the following statements:

(E1) There exists an $x \in \mathsf{X}$ such that $v_\sigma(x) = v^*(x)$

(E2) There exists a distribution ρ on X such that

$$\int v_\sigma(x)\rho(\mathrm{d}x) = \int v^*(x)\rho(\mathrm{d}x)$$

(E3) σ is an optimal policy

We seek conditions under which (E1)–(E3) are equivalent

Irreducibility (Banach lattice setting)

Def. Let

- E be a Banach lattice
- E' be the dual space
- E_+ and E'_+ be the respective positive cones

A positive linear operator K on E is called **irreducible** if

- for all nonzero $f \in E_+$
- and all nonzero $\mu \in E'_+$

there exists an $n \in \mathbb{N}$ with $\langle \mu, K^n f \rangle > 0$

Theorem 1: Global Optimality and Irreducibility

If P_σ is irreducible, then $(\mathrm{E1})$–$(\mathrm{E3})$ are equivalent

Implications:

- Maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ **always** produces an optimal σ
- Maximizing $v_\sigma(x)$ at fixed x **always** produces an optimal policy

Intuition:

- Irreducibility under σ supplies the mixing

Theorem 1: Global Optimality and Irreducibility

If P_σ is irreducible, then $(\mathrm{E1})$–$(\mathrm{E3})$ are equivalent

Implications:

- Maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ **always** produces an optimal σ
- Maximizing $v_\sigma(x)$ at fixed x **always** produces an optimal policy

Intuition:

- Irreducibility under σ supplies the mixing

What if irreducibility fails?

Do weaker conditions work in some settings?

Let K be any Markov kernel on X

Def. A point $y \in \mathrm{X}$ is called **reachable** from $x \in \mathrm{X}$ if, for each open neighborhood G of y, there exists an $n \in \mathbb{N}$ with $K^n(x, G) > 0$.

Def. K is called **open set irreducible** if every $y \in \mathrm{X}$ is reachable from every $x \in \mathrm{X}$

What if irreducibility fails?

Do weaker conditions work in some settings?

Let K be any Markov kernel on X

Def. A point $y \in$ X is called **reachable** from $x \in$ X if, for each open neighborhood G of y, there exists an $n \in \mathbb{N}$ with $K^n(x, G) > 0$.

Def. K is called **open set irreducible** if every $y \in$ X is reachable from every $x \in$ X

Theorem 2: Global Optimality and Open Set Irreducibility

If P_σ is open set irreducible and the policy σ is continuous, then (E1)–(E3) are equivalent

Implications:

- Maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ **always** produces an optimal σ
- Maximizing $v_\sigma(x)$ at fixed x **always** produces an optimal policy

Intuition:

- Irreducibility under σ supplies the mixing

Theorem 2: Global Optimality and Open Set Irreducibility

If P_σ is open set irreducible and the policy σ is continuous, then (E1)–(E3) are equivalent

Implications:

- Maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ **always** produces an optimal σ
- Maximizing $v_\sigma(x)$ at fixed x **always** produces an optimal policy

Intuition:

- Irreducibility under σ supplies the mixing

Theorem 2: Global Optimality and Open Set Irreducibility

If P_σ is open set irreducible and the policy σ is continuous, then (E1)–(E3) are equivalent

Implications:

- Maximizing $F(\sigma) = \int v_\sigma(x)\rho(\mathrm{d}x)$ **always** produces an optimal σ
- Maximizing $v_\sigma(x)$ at fixed x **always** produces an optimal policy

Intuition:

- Irreducibility under σ supplies the mixing

Example: optimal saving

Wealth obeys

$$W_{t+1} = R_{t+1}(W_t - C_t) + Y_{t+1}$$

where

- (Y_t) and (R_t) are IID with distributions φ and ψ
- $\mathsf{X} = \mathsf{A} = \mathbb{R}_+$
- a policy is a map σ with $0 \leqslant \sigma(w) \leqslant w$ for all $w \in \mathbb{R}_+$

Assume u is bounded, increasing, continuous, and strictly concave

Standard optimality results hold

Proposition 2

Under the stated conditions

1. The value function satisfies the Bellman equation
2. The optimal savings problem has a unique optimal policy σ^*
3. The policy σ^* is continuous
4. For all $w \in \mathbb{R}_+$,

$$\sigma^*(w) \in \operatorname*{arg\,max}_{0 \leqslant c \leqslant w} \left\{ u(c) + \beta \int v(r(w-c)+y)\psi(\mathrm{d}r)\varphi(\mathrm{d}y) \right\}$$

5. Value function iteration converges

To apply our theory we need sufficient mixing

Proposition 3

If ψ and φ have full support on $\mathbb{R}_+$, then

$$P_\sigma \text{ is open set irreducible for all } \sigma \in \Sigma$$

We **now assume** that ψ, φ have full support on $\mathbb{R}_+$

As a result,

$$\text{local optimality} \implies \text{global optimality}$$

To apply our theory we need sufficient mixing

Proposition 3

If ψ and φ have full support on $\mathbb{R}_+$, then

$$P_\sigma \text{ is open set irreducible for all } \sigma \in \Sigma$$

We **now assume** that ψ, φ have full support on $\mathbb{R}_+$

As a result,

$$\text{local optimality} \implies \text{global optimality}$$

Experiment

1. Fix $\bar{w} \in \mathbb{R}_+$
2. Compute $\hat{\sigma}$ by solving $\max_{\sigma \in \Sigma} v_\sigma(\bar{w})$ via policy gradient ascent
3. Compute σ^* via a traditional DP algorithm
4. Compare σ^* and $\hat{\sigma}$ on all of $\mathbb{R}_+$
5. Compare v^* and $v_{\hat{\sigma}}$ on all of $\mathbb{R}_+$

Experiment

1. Fix $\bar{w} \in \mathbb{R}_+$
2. Compute $\hat{\sigma}$ by solving $\max_{\sigma \in \Sigma} v_\sigma(\bar{w})$ via policy gradient ascent
3. Compute σ^* via a traditional DP algorithm
4. Compare σ^* and $\hat{\sigma}$ on all of $\mathbb{R}_+$
5. Compare v^* and $v_{\hat{\sigma}}$ on all of $\mathbb{R}_+$

Our results predict that $\hat{\sigma}$ is optimal

As a result, we have

$$v_{\hat{\sigma}} = v^* \text{ on all of } \mathbb{R}_+$$

Since the optimal policy is unique, we also predict

$$\hat{\sigma} = \sigma^* \text{ on all of } \mathbb{R}_+$$

Our results predict that $\hat{\sigma}$ is optimal

As a result, we have

$$v_{\hat{\sigma}} = v^* \text{ on all of } \mathbb{R}_+$$

Since the optimal policy is unique, we also predict

$$\hat{\sigma} = \sigma^* \text{ on all of } \mathbb{R}_+$$

In practice

1. Replace Σ with parametrized policies
   - $\Sigma_\theta := \{\sigma(\cdot, \theta) : \theta \in \Theta\}$
   - each $\sigma(\cdot, \theta)$ is an ANN

2. Set
$$F(\theta) := v_{\sigma(\cdot,\theta)}(\bar{w})$$

3. Apply gradient ascent: $\theta_{n+1} = \theta_n + \lambda_n \nabla F(\theta_n)$

In the computation, we need to evaluate

$$F(\theta) = \mathbb{E}_{\bar{w}} \sum_{t=0}^{\infty} \beta^t u(\sigma(W_t, \theta))$$

- $W_0 = \bar{w}$ and $W_{t+1} = R_{t+1}(W_t - \sigma(W_t, \theta)) + Y_{t+1}$

We approximate $F(\theta)$ by MC:

$$\hat{F}(\theta) := \frac{1}{N} \sum_{i=1}^{N} \sum_{t=0}^{T-1} \beta^t u(\sigma(W_t^i, \theta))$$

For gradient ascent, we compute $\nabla \hat{F}$ using autodiff

Let $\hat{\sigma} = \hat{\sigma}(\cdot, \theta)$ be the policy returned by the algorithm

Now evaluate $v_{\hat{\sigma}}$ everywhere on $\mathbb{R}_+$

1. Fix an initial guess v_0
2. Iterate to convergence with

$$(T_{\hat{\sigma}}\, v)(w) = u(\hat{\sigma}(w)) + \beta \int v(r(w - \hat{\sigma}(w)) + y)\psi(\mathrm{d}r)\varphi(\mathrm{d}y)$$

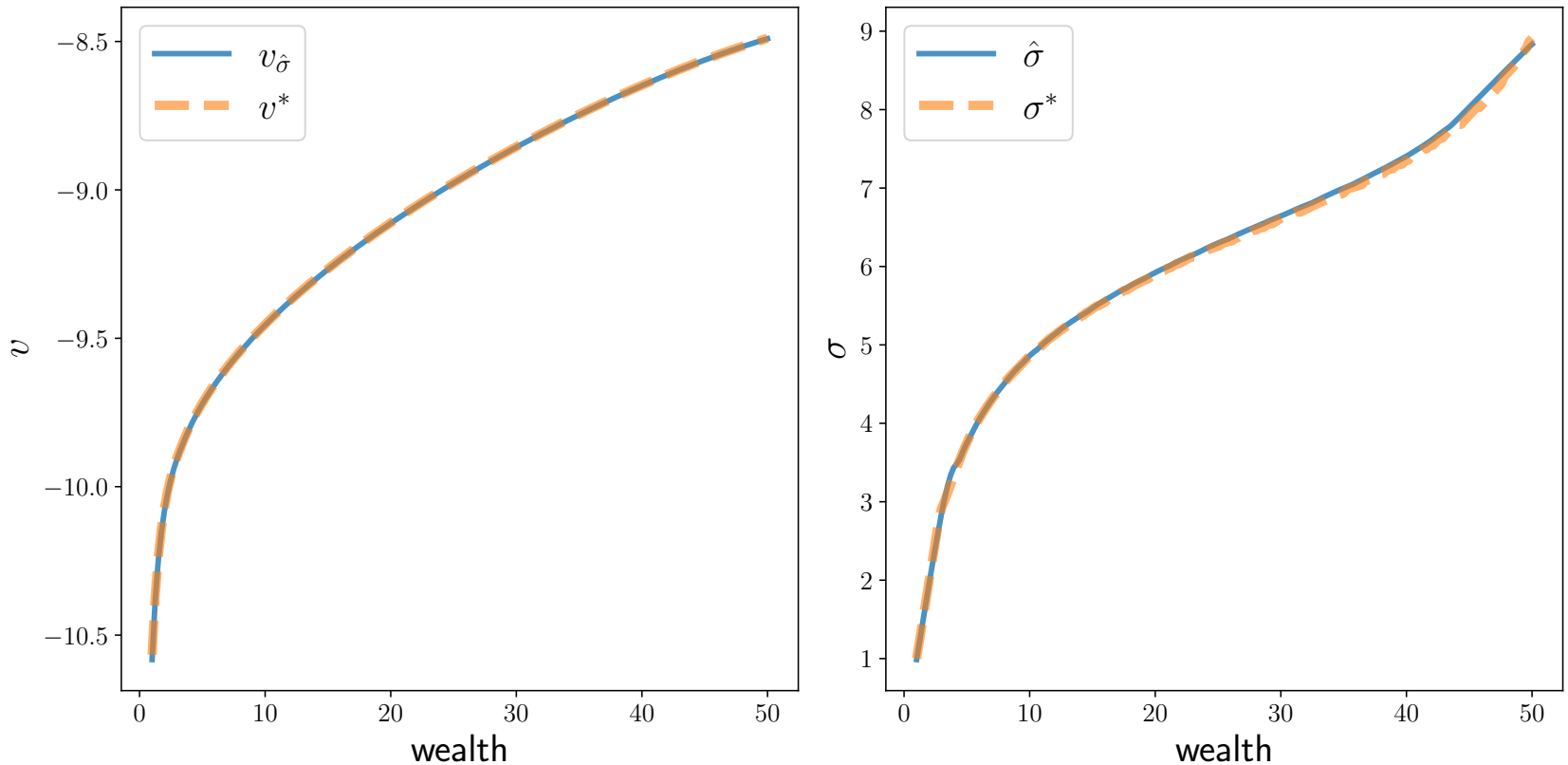


Figure 1: $v_{\hat{\sigma}}$ and $\hat{\sigma}$ with $\bar{w} = 1$ vs OPI solutions

Breaking irreducibility

Now suppose that returns and labor income are **bounded**

- Assume ψ and φ are uniform on intervals

$$[\underline{r}, \overline{r}] = [0.5, 0.8] \quad \text{and} \quad [\underline{y}, \overline{y}] = [1, 8]$$

In this setting, irreducibility is lost...

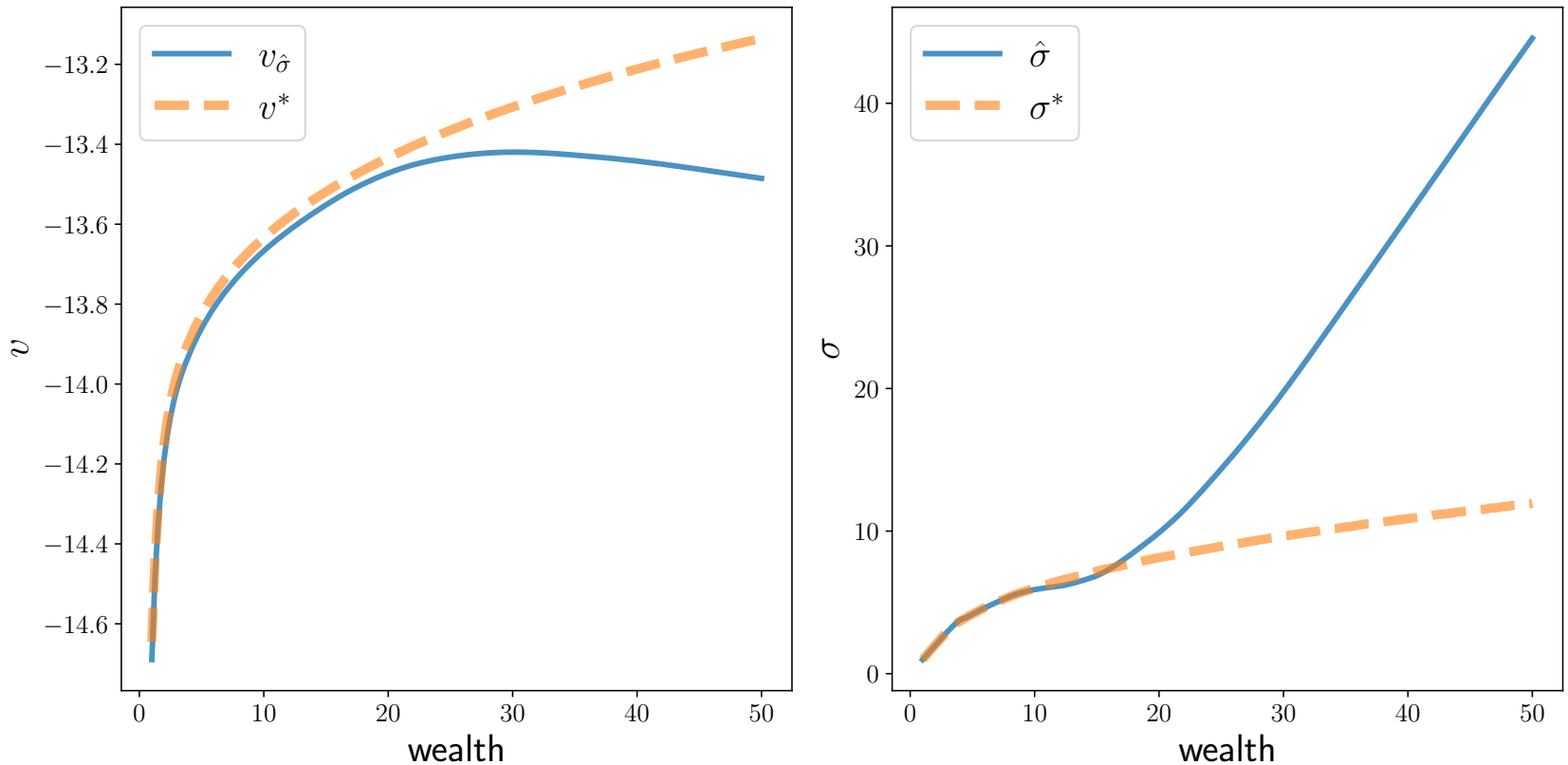


Figure 2: $\hat{v}_\sigma$ and $\hat{\sigma}$ with $\bar{w} = 1$ against the OPI solutions.

Questions

What about

- "Partial" irreducibility?
- Unbounded rewards?
- State-dependent discounting?
- Recursive preference models?
- Continuous time MDPs?
- Dynamic games?