Dynamic Programming

Chapter 2: Operators and Fixed Points

Thomas J. Sargent and John Stachurski

2024

Summary

- Conjugate maps
- Convergence rates
- Newton's method
- Partial orders
- Order-preserving maps
- Fixed points and order
- Linear operators

Conjugate Maps

Suppose we are concerned with the dynamics induced by a self-map T on $\mathbb{R}^n$

- does a unique fixed point of T exist?
- do iterates of T always converge to a fixed point?

Option A: Apply fixed point theory to T

Option B:

1. transform T into a "simpler" operator $\hat{T}$
2. apply fixed point theory to $\hat{T}$
3. translate properties we discover about $\hat{T}$ back to T

To implement Option B we need the following definitions:

A **dynamical system** is a pair (U, T), where

- U is a subset of $\mathbb{R}^n$ and
- T is a self-map on U

Dynamical systems (U, T) and $(\hat{U}, \hat{T})$ are called **conjugate** under Φ if

1. Φ is a bijection from U into $\hat{U}$ and
2. $T = \Phi^{-1} \circ \hat{T} \circ \Phi$ on U

The condition $T = \Phi^{-1} \circ \hat{T} \circ \Phi$ can be understood as follows:

Shifting a point $u \in U$ to Tu via T is equivalent to

1. shifting u from U to $\hat{U}$ via $\hat{u} = \Phi u$
2. applying $\hat{T}$, and
3. shifting the result back to U via Φ^{-1}:

$$\begin{tikzcd} u \arrow[r, "T"] \arrow[d, "\Phi"] & Tu \\ \hat{u} \arrow[r, "\hat{T}"] & \hat{T}\hat{u} \arrow[u, "\Phi^{-1}"'] \end{tikzcd}$$

Ex. (Log-linearization) Fix $A > 0$, $\alpha \in \mathbb{R}$ and suppose

- $U := (0, \infty)$ and $Tu = Au^\alpha$
- $\hat{U} := \mathbb{R}$ and $\hat{T}\hat{u} = \ln A + \alpha \hat{u}$

Show that (U, T) and $(\hat{U}, \hat{T})$ are conjugate under $\Phi := \ln$

Proof: The condition $T = \Phi^{-1} \circ \hat{T} \circ \Phi$ is equivalent to

$$\Phi \circ T = \hat{T} \circ \Phi$$

This holds because, for $u \in \mathbb{R}$,

- $\Phi \, T \, u = \ln A + \alpha \ln u$
- $\hat{T} \, \Phi \, u = \ln A + \alpha \ln u$

Ex. (Log-linearization) Fix $A > 0$, $\alpha \in \mathbb{R}$ and suppose

- $U := (0, \infty)$ and $Tu = Au^\alpha$
- $\hat{U} := \mathbb{R}$ and $\hat{T}\hat{u} = \ln A + \alpha\hat{u}$

Show that (U, T) and $(\hat{U}, \hat{T})$ are conjugate under $\Phi := \ln$

Proof: The condition $T = \Phi^{-1} \circ \hat{T} \circ \Phi$ is equivalent to

$$\Phi \circ T = \hat{T} \circ \Phi$$

This holds because, for $u \in \mathbb{R}$,

- $\Phi\, T\, u = \ln A + \alpha \ln u$
- $\hat{T}\, \Phi\, u = \ln A + \alpha \ln u$

Ex. Let (U, T) and $(\hat{U}, \hat{T})$ be conjugate under Φ

Show that the following statements are equivalent

1. $u \in U$ is a fixed point of T on U
2. $\Phi u \in \hat{U}$ is a fixed point of $\hat{T}$ on $\hat{U}$

Proof: Let (U, T) and $(\hat{U}, \hat{T})$ be as stated

- then $\hat{T} \circ \Phi = \Phi \circ T$

The claimed equivalence holds because

$$Tu = u \iff \Phi T u = \Phi u \iff \hat{T} \Phi u = \Phi u$$

Ex. Let (U, T) and $(\hat{U}, \hat{T})$ be conjugate under Φ

Show that the following statements are equivalent

1. $u \in U$ is a fixed point of T on U
2. $\Phi u \in \hat{U}$ is a fixed point of $\hat{T}$ on $\hat{U}$

Proof: Let (U, T) and $(\hat{U}, \hat{T})$ be as stated

- then $\hat{T} \circ \Phi = \Phi \circ T$

The claimed equivalence holds because

$$Tu = u \iff \Phi T u = \Phi u \iff \hat{T} \Phi u = \Phi u$$

Topological Conjugacy

Let U and $\hat{U}$ be two subsets of $\mathbb{R}^n$

$\Phi\colon U \to \hat{U}$ is called a **homeomorphism** if it is a continuous bijection and Φ^{-1} is also continuous

Example. $\Phi = \ln$ is a homeomorphism from $(0, \infty) \to \mathbb{R}$ with

$$\Phi^{-1} = \exp$$

Example. Let A be an $n \times n$ matrix understood as a map $u \mapsto Au$

$$A\colon \mathbb{R}^n \to \mathbb{R}^n \text{ is a homeomorphism } \iff A \text{ is nonsingular}$$

Let

- (U, T) and $(\hat{U}, \hat{T})$ be two dynamical systems
- Φ be a map from U to $\hat{U}$

We call (U, T) and $(\hat{U}, \hat{T})$ **topologically conjugate** under Φ if

1. (U, T) and $(\hat{U}, \hat{T})$ are conjugate under Φ and
2. Φ is a homeomorphism

Example. An $n \times n$ matrix A is called **diagonalizable** if $\exists$ a diagonal matrix D and a nonsingular matrix P such that

$$A = P^{-1} D P$$

- D and P can be complex-valued
- we view A as a self-map on $\mathbb{R}^n$
- and D and a self-map on $\mathbb{C}^n$

Note that P and P^{-1} are continuous bijections

- linear maps between finite-dimensional spaces are continuous

Hence $(A, \mathbb{R}^n)$ and $(D, \mathbb{C}^n)$ are topologically conjugate

Ex. Let (U, T) and $(\hat{U}, \hat{T})$ be topologically conjugate under Φ

Given $u, u^* \in U$, show that

$$T^k u \to u^* \quad \iff \quad \hat{T}^k \Phi u \to \Phi u^*$$

Proof: From $\hat{T} = \Phi \circ T \circ \Phi^{-1}$ we can show that

$$\hat{T}^k = \Phi \circ T^k \circ \Phi^{-1} \quad \text{for all } k \in \mathbb{N}$$

Hence, using continuity of Φ and Φ^{-1},

$$T^k u \to u^* \iff \Phi T^k u \to \Phi u^* \iff \hat{T}^k \Phi u \to \Phi u^*$$

Ex. Let (U, T) and $(\hat{U}, \hat{T})$ be topologically conjugate under Φ

Given $u, u^* \in U$, show that

$$T^k u \to u^* \quad \iff \quad \hat{T}^k \Phi u \to \Phi u^*$$

Proof: From $\hat{T} = \Phi \circ T \circ \Phi^{-1}$ we can show that

$$\hat{T}^k = \Phi \circ T^k \circ \Phi^{-1} \quad \text{for all } k \in \mathbb{N}$$

Hence, using continuity of Φ and Φ^{-1},

$$T^k u \to u^* \iff \Phi T^k u \to \Phi u^* \iff \hat{T}^k \Phi u \to \Phi u^*$$

Let (U, T) and $(\hat{U}, \hat{T})$ be two dynamical systems

Proposition. If (U, T) and $(\hat{U}, \hat{T})$ are topologically conjugate, then

$$(U, T) \text{ is globally stable } \iff (\hat{U}, \hat{T}) \text{ is globally stable}$$

Moreover, if one and hence both are globally stable, then the unique fixed points $u^* \in U$ and $\hat{u}^* \in \hat{U}$ satisfy

$$\hat{u}^* = \Phi u^*$$

Ex. Prove this

Local Stability

Let U be a subset of $\mathbb{R}^n$ and let T be a self-map on U

A fixed point u^* of T in U is called **locally stable** for T if $\exists$ an open set $O \subset U$ such that

$$u^* \in O \text{ and } T^k u \to u^* \text{ as } k \to \infty \text{ for every } u \in O$$

Ex. Let (U, T) and $(\hat{U}, \hat{T})$ be topologically conjugate and let u^* be a fixed point of T in U

Show that

$$u^* \text{ is locally stable for } T \iff \Phi u^* \text{ is locally stable for } \hat{T}$$

Derivative tests

One way to verify local stability of a one-dimensional map g at fixed point x^* is to show that $|g'(x^*)| < 1$

Intuition: The first-order linear approximation of g near x^* is

$$\begin{aligned}\hat{g}(x) &:= g(x^*) + g'(x^*)(x - x^*) \\ &= x^* + g'(x^*)(x - x^*)\end{aligned}$$

When $|g'(x^*)| < 1$, the map $\hat{g}$ is a contraction of modulus $|g'(x^*)|$

Moreover, $\hat{g}$ and g have "the same" dynamics near x^*

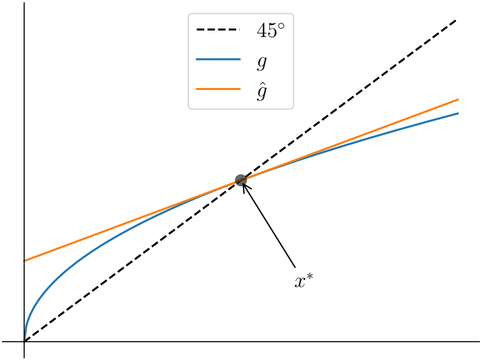


Figure: Local stability when $|g'(x^*)| < 1$

Let's formalize this argument and generalize to vector space

Take T to be a continuously differentiable self-map on $U \subset \mathbb{R}^n$ with fixed point u^*

Recall that the **Jacobian** of T at $u \in U$ is

$$J_T(u) := \begin{pmatrix} \frac{\partial T_1}{\partial u_1}(u) & \cdots & \frac{\partial T_1}{\partial u_n}(u) \\ & \cdots & \\ \frac{\partial T_n}{\partial u_1}(u) & \cdots & \frac{\partial T_n}{\partial u_n}(u) \end{pmatrix} \quad \text{where} \quad Tu = \begin{pmatrix} T_1 u \\ \vdots \\ T_n u \end{pmatrix}$$

We set $\hat{T}$ to be the first-order approximation to T at u^*:

$$\hat{T}u = u^* + J_T(u^*)(u - u^*) \qquad (u \in U)$$

Theorem. (Hartman–Grobman) If $J_T(u^*)$

1. is nonsingular and
2. has no eigenvalues on the unit circle in $\mathbb{C}$,

then $\exists$ an open neighborhood O of u^* such that (O, T) and $(O, \hat{T})$ are topologically conjugate

In particular,

$$u^* \text{ locally stable for } T \text{ whenever } \hat{T} \text{ globally stable on } \mathbb{R}^n$$

Recall that

$$\hat{T}u = u^* + J_T(u^*)(u - u^*) \qquad (u \in U)$$

By the Neumann series lemma,

$$\rho(J_T(u^*)) < 1 \implies \hat{T} \text{ is globally stable on } \mathbb{R}^n$$

Corollary. Under the conditions of the Hartman–Grobman theorem,

$$\rho(J_T(u^*)) < 1 \quad \implies \quad u^* \text{ is locally stable for } T$$

Convergence Rates

Fix norm $\|\cdot\|$ on $\mathbb{R}^n$

In what follows

1. $(u_k)_{k\geqslant 0} \subset \mathbb{R}^n$ converges to $u^* \in \mathbb{R}^n$
2. $e_k := \|u_k - u^*\|$

In particular, we have

$$e_k \to 0 \qquad (k \to \infty)$$

We wish to quantify the rate of convergence

We say $(u_k)_{k \geqslant 0}$ converges to u^* **at rate at least q** if

1. $q \geqslant 1$
2. for some $\beta \in (0, \infty)$ and $N \in \mathbb{N}$,

$$e_{k+1} \leqslant \beta e_k^q \quad \text{for all } k \geqslant N$$

We say that convergence occurs **at rate q** if, in addition,

$$\limsup_{k \to \infty} \frac{e_{k+1}}{e_k^q} = \beta$$

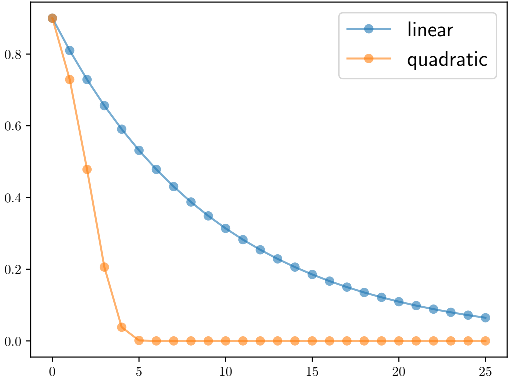
- If $q = 2$ we say that convergence is (at least) **quadratic**
- If $q = 1$ and $\beta < 1$, we say convergence is (at least) **linear**

Example. With $|\beta| \in (0,1)$ and $q \geqslant 1$, the real sequence

$$x_{k+1} = \beta x_k^q, \qquad |x_0| < 1$$

converges to 0 at rate q because

$$\begin{aligned} \limsup_{k\to\infty} \frac{e_{k+1}}{e_k^q} &= \limsup_{k\to\infty} \frac{|x_{k+1}|}{|x_k^q|} \\ &= \limsup_{k\to\infty} \frac{|\beta||x_k^q|}{|x_k^q|} \\ &= |\beta| \end{aligned}$$


linear
quadratic
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25

Example. Suppose u_k converges to u^* quadratically, so that

$$\limsup_{k\to\infty} \frac{e_{k+1}}{e_k^2} = \beta$$

Suppose also that β is not large

If, say, $e_k = 10^{-5}$, then

$$e_{k+1} \approx \beta 10^{-10} \approx 10^{-10}$$

Number of accurate digits roughly doubles at each step

Example. Let

- T be a contraction of modulus λ on closed set $U \subset \mathbb{R}^n$
- u^* be the unique fixed point of T in U
- u be fixed and $u_k := T^k u$

Then (u_k) converges at least linearly to u^*, since

$$\begin{aligned} e_{k+1} &= \|u_{k+1} - u^*\| \\ &= \|Tu_k - Tu^*\| \\ &\leqslant \lambda \|u_k - u^*\| \\ &= \lambda e_k \end{aligned}$$

Rule of thumb: Successive approximation often converges linearly

The next exercise provides some evidence

Ex. Let $T \colon U \to U$ be smooth on open interval U in $\mathbb{R}$

Suppose that

- T has a fixed point $u^* \in U$ and
- $u_k := T^k u_0$ converges to u^* as $k \to \infty$

Prove that the rate of convergence is linear whenever

$$0 < |T' u^*| < 1$$

Hint: Taylor expansion yields a $v_k \in (u_k, u^*)$ such that

$$Tu_k = u^* + T'u^*(u_k - u^*) + \frac{T''v_k}{2}(u_k - u^*)^2$$

Proof: Since $u_{k+1} = Tu_k$, we have

$$\frac{u_{k+1} - u^*}{u_k - u^*} = T'u^* + \frac{T''v_k}{2}(u_k - u^*)$$

$$\therefore \quad \limsup_{k\to\infty} \frac{e_{k+1}}{e_k} = |T'u^*|$$

Hint: Taylor expansion yields a $v_k \in (u_k, u^*)$ such that

$$Tu_k = u^* + T'u^*(u_k - u^*) + \frac{T''v_k}{2}(u_k - u^*)^2$$

<u>Proof</u>: Since $u_{k+1} = Tu_k$, we have

$$\frac{u_{k+1} - u^*}{u_k - u^*} = T'u^* + \frac{T''v_k}{2}(u_k - u^*)$$

$$\therefore \quad \limsup_{k\to\infty} \frac{e_{k+1}}{e_k} = |T'u^*|$$

Newton's Method

Successive approximation is typically linear

Faster algorithms can often be obtained for smooth functions

Strategy: leverage the information provided by gradients

One important gradient-based technique is Newton's method

Suppose that

- T is a differentiable self-map on an open set $U \subset \mathbb{R}^n$
- our aim is to find a fixed point of T

Our plan: start with a guess u_0 and then update it to u_1

To do this we

1. construct the first-order approximation $\hat{T}$ of T around u_0
2. obtain the fixed point of $\hat{T}$ (exactly, since $\hat{T}$ is linear)

We take this new point u_1 and repeat

The next figure shows u_0, u_1 when

- $n = 1$
- $Tu = 1 + u/(u+1)$

- $u_0 = 0.5$
- $\hat{T}$ is the first order approximation at u_0
- u_1 is the fixed point of $\hat{T}$

Note u_1 is closer to the fixed point of T than u_0, as desired

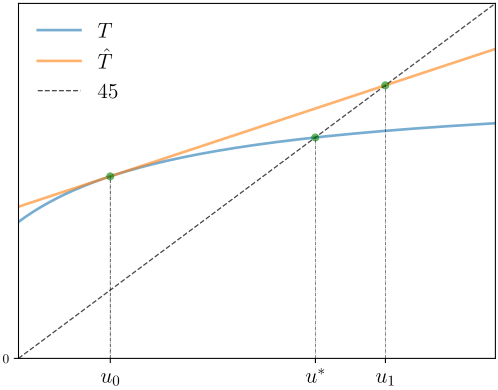


Figure: First step of Newton's method: from u_0 to u_1

Shifting to the n-dimensional case, let

- $J_T(u_0) =$ Jacobian of T at u_0 and $I = n \times n$ identity

The first order approximation at u_0 is

$$\hat{T}u := Tu_0 + J_T(u_0)(u - u_0)$$

We seek the u_1 such that $\hat{T}u_1 = u_1$, or

$$u_1 = Tu_0 + J_T(u_0)(u_1 - u_0)$$

Solving for u_1 gives

$$u_1 = (I - J_T(u_0))^{-1}(Tu_0 - J_T(u_0)u_0)$$

Now repeat starting at u_1, etc.

More generally, define

$$u_{k+1} = Qu_k \qquad (k \geqslant 0)$$

where

$$Qu := (I - J_T(u))^{-1}(Tu - J_T(u)u) \qquad k = 0, 1, \ldots$$

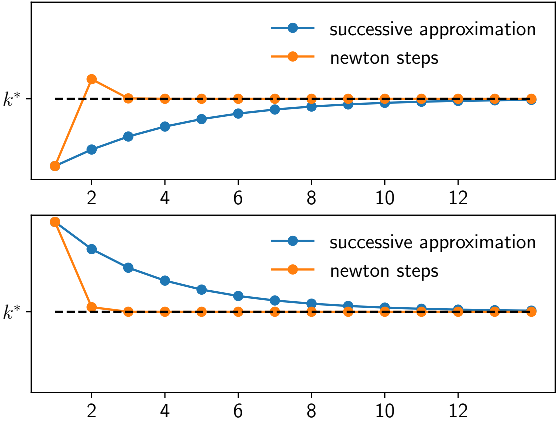
Newton's fixed point method: apply successive approximation to Q

The next figure shows an application to computing the fixed point of the Solow–Swan model

We use

- both Newton's method and successive approximation
- same initial conditions

Both sequences converge but the Newton sequences converge faster


successive approximation
newton steps
k^*
2
4
6
8
10
12
successive approximation
newton steps
k^*
2
4
6
8
10
12

Fast rates of convergence can be confirmed theoretically for the Newton scheme

Under mild conditions, $\exists$ a neighbourhood of the fixed point within which the Newton iterates converge quadratically

- See the text for references

This fast rate of convergence will be significant when we study dynamic programming algorithms

- An algorithm called Howard policy iteration is a version of Newton's method

Parallelization

Successive approx. is highly serial: cannot compute $T^{k+1}u$ until $T^k u$ is available

- slow convergence
- many sequential steps

Newton's method is also serial — we are just iterating with a different map — but

- fewer steps
- each one is more computationally intensive

Hence well suited to parallelization

Automatic differentiation

Many software platforms now offer automatic differentiation

- simular to symbolic (exact) differentiation
- but more numerically efficient
- also more efficient/stable than numerical differentiation

Automatic differentiation can be used for computing Jacobians in the Newton step

Combines well with parallelization

Order

Let's review the fundamentals of **order theory**

One of the foundational subjects of maths, on par with

- algebra
- geometry
- topology, etc.

But not commonly taught in foundational math courses

Why?

Rarely used in

- physics
- chemistry
- biology, etc.

Math courses are biased toward these subjects!

But not commonly taught in foundational math courses

Why?

Rarely used in

- physics
- chemistry
- biology, etc.

Math courses are biased toward these subjects!

But very important for econ / OR / finance

Examples.

- Does consumer X <u>prefer</u> good A or good B?
- Is welfare <u>greater</u> under policy A or policy B?
- Does R & D <u>increase</u> profits?
- How can firm Y <u>minimize</u> costs?

In these lectures, we need order for

- studying optimality
- fixed point results

Partial orders

Let P be a nonempty set

A **partial order** on a P is a binary relation $\preceq$ on $P \times P$ satisfying, for any p, q, r in P,

$p \preceq p$,

$p \preceq q$ and $q \preceq p$ implies $p = q$ and

$p \preceq q$ and $q \preceq r$ implies $p \preceq r$

(Reflexivity, antisymmetry, transitivity)

We call $(P, \preceq)$ a **partially ordered set**

- $P := (P, \preceq)$ when $\preceq$ understood

Ex.

1. Show that the usual order $\leqslant$ on $\mathbb{R}$ is a partial order on $\mathbb{R}$
2. Given set M, show that $\subset$ is a partial order on $\wp(M)$

Proof for 2: Clearly, for all $A, B, C \subset M$,

- $A \subset A$ holds
- $A \subset B$ and $B \subset A$ implies $A = B$
- $A \subset B$ and $B \subset C$ implies $A \subset C$

Ex.

1. Show that the usual order $\leqslant$ on $\mathbb{R}$ is a partial order on $\mathbb{R}$
2. Given set M, show that $\subset$ is a partial order on $\wp(M)$

Proof for 2: Clearly, for all $A, B, C \subset M$,

- $A \subset A$ holds
- $A \subset B$ and $B \subset A$ implies $A = B$
- $A \subset B$ and $B \subset C$ implies $A \subset C$

A partial order $\preceq$ on P is called a **total order** if

$$\text{either } p \preceq q \text{ or } q \preceq p \text{ for all } p, q \in P$$

Example. $\leqslant$ is a total order on $\mathbb{R}$

Ex. Prove: $\subset$ is not a total order on $\wp(M)$ when $|M| > 1$

Proof: If $|M| \geqslant 2$, then $\exists$ nonempty $A, B \subset M$ with $A \cup B = \emptyset$

But then $A \subset B$ and $B \subset A$ both fail

A partial order $\preceq$ on P is called a **total order** if

$$\text{either } p \preceq q \text{ or } q \preceq p \text{ for all } p, q \in P$$

Example. $\leqslant$ is a total order on $\mathbb{R}$

Ex. Prove: $\subset$ is not a total order on $\wp(M)$ when $|M| > 1$

Proof: If $|M| \geqslant 2$, then $\exists$ nonempty $A, B \subset M$ with $A \cup B = \emptyset$

But then $A \subset B$ and $B \subset A$ both fail

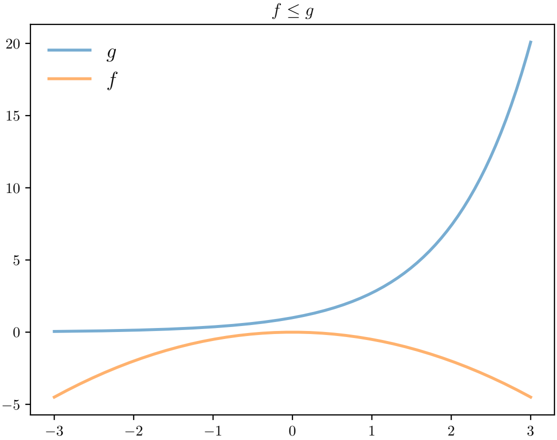
Pointwise Orders

Let

- X be any set
- $\mathbb{R}^{\mathsf{X}}$ be all $f\colon \mathsf{X} \to \mathbb{R}$

The **pointwise order** over $\mathbb{R}^{\mathsf{X}}$ is written as $\leqslant$ and defined via

$$f \leqslant g \iff f(x) \leqslant g(x) \text{ for all } x \in \mathsf{X}$$


$f \leq g$
g
f
20
15
10
5
0
−5
−3
−2
−1
0
1
2
3

Ex. Show $\leqslant$ is a partial order on $\mathbb{R}^{\mathsf{X}}$

Proof:

Let's just check antisymmetry

Fix $f, g \in \mathbb{R}^{\mathsf{X}}$ and suppose $f \leqslant g$ and $g \leqslant f$

Pick any $x \in \mathsf{X}$

By definition, $f(x) \leqslant g(x)$ and $g(x) \leqslant f(x)$

Therefore, $f(x) = g(x)$

Since x was arbitrary, we have $f = g$

Ex. Show $\leqslant$ is a partial order on $\mathbb{R}^{\mathsf{X}}$

Proof:

Let's just check antisymmetry

Fix $f, g \in \mathbb{R}^{\mathsf{X}}$ and suppose $f \leqslant g$ and $g \leqslant f$

Pick any $x \in \mathsf{X}$

By definition, $f(x) \leqslant g(x)$ and $g(x) \leqslant f(x)$

Therefore, $f(x) = g(x)$

Since x was arbitrary, we have $f = g$

Let's define the **pointwise order for matrices**

Let $\mathbb{M}^{n\times k} :=$ all $n \times k$ matrices

For $A = (a_{ij})$ and $B = (b_{ij})$ in $\mathbb{M}^{n\times k}$, we set

$$A \leqslant B \iff a_{ij} \leqslant b_{ij} \text{ for all } i, j$$

Example.

$$\begin{pmatrix} 1 & 2 \\ -2 & 0 \end{pmatrix} \leqslant \begin{pmatrix} 10 & 20 \\ 0 & 10 \end{pmatrix}$$

Ex. Show that $\leqslant$ is a partial order on $\mathbb{M}^{n\times k}$

Special case: **pointwise order for vectors**

Recall $[n] := \{1, \dots, n\}$

For $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, we write

$$x \leqslant y \quad \iff \quad x_i \leqslant y_i \text{ for all } i \in [n]$$

Pointwise order $\leqslant$ on $\mathbb{R}^2$:

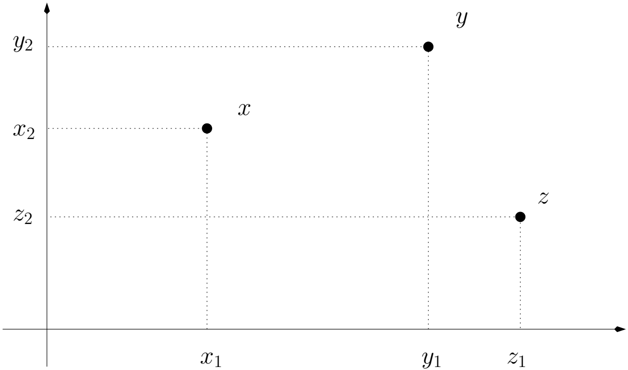


Figure: $x \leqslant y$ but neither x, z nor y, z are comparable

Ex. Prove: for $a, b \in \mathbb{R}^n$ and sequence (x_k) in $\mathbb{R}^n$, we have

$$a \leqslant x_k \leqslant b \text{ for all } k \in \mathbb{N} \text{ and } x_k \to x \text{ implies } a \leqslant x \leqslant b$$

Proof: Fix $i \in [n]$

Let a^i be the i-th element of a, etc.

It suffices to show that

$$a^i \leqslant x^i \leqslant b^i \tag{1}$$

Note $x_k \to x$ implies $x_k^i \to x^i$

Moreover, $a^i \leqslant x_k^i \leqslant b^i$ for all k

Weak inequalities in $\mathbb{R}$ are preserved under limits, so (1) holds

Ex. Prove: for $a, b \in \mathbb{R}^n$ and sequence (x_k) in $\mathbb{R}^n$, we have

$$a \leqslant x_k \leqslant b \text{ for all } k \in \mathbb{N} \text{ and } x_k \to x \text{ implies } a \leqslant x \leqslant b$$

<u>Proof</u>: Fix $i \in [n]$

Let a^i be the i-th element of a, etc.

It suffices to show that

$$a^i \leqslant x^i \leqslant b^i \tag{1}$$

Note $x_k \to x$ implies $x_k^i \to x^i$

Moreover, $a^i \leqslant x_k^i \leqslant b^i$ for all k

Weak inequalities in $\mathbb{R}$ are preserved under limits, so (1) holds

In other words, the pointwise order $\leqslant$ is preserved under limits

As a result, these sets are **closed**

- $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : 0 \leqslant x\}$
- $[a, b] := \{x \in \mathbb{R}^n : a \leqslant x \leqslant b\}$
- etc.

A key connection between order and topology!

Ex. Prove: If B is $m \times k$ and $B \geqslant 0$, then

$$|Bx| \leqslant B|x| \text{ for all } k \times 1 \text{ column vectors } x$$

Proof: Fix $B \in \mathbb{M}^{m \times k}$ with $b_{ij} \geqslant 0$ for all i, j

Fix $i \in [m]$ and $x \in \mathbb{R}^k$

By the triangle inequality, we have $|\sum_j b_{ij} x_j| \leqslant \sum_j b_{ij} |x_j|$

Stacking these inequalities yields

$$|Bx| \leqslant B|x|$$

Ex. Prove: If B is $m \times k$ and $B \geqslant 0$, then

$$|Bx| \leqslant B|x| \text{ for all } k \times 1 \text{ column vectors } x$$

Proof: Fix $B \in \mathbb{M}^{m \times k}$ with $b_{ij} \geqslant 0$ for all i, j

Fix $i \in [m]$ and $x \in \mathbb{R}^k$

By the triangle inequality, we have $|\sum_j b_{ij} x_j| \leqslant \sum_j b_{ij} |x_j|$

Stacking these inequalities yields

$$|Bx| \leqslant B|x|$$

Lemma. If X is finite and $f, g, h \in \mathbb{R}^{\mathsf{X}}$, then

1. $|f + g| \leqslant |f| + |g|$
2. $(f \wedge g) + h = (f + h) \wedge (g + h)$
3. $(f \vee g) + h = (f + h) \vee (g + h)$
4. $(f \vee g) \wedge h = (f \wedge h) \vee (g \wedge h)$
5. $(f \wedge g) \vee h = (f \vee h) \wedge (g \vee h)$
6. $|f \wedge h - g \wedge h| \leqslant |f - g|$
7. $|f \vee h - g \vee h| \leqslant |f - g|$

Also, if $f, g, h \in \mathbb{R}_+^{\mathsf{X}}$, then

$$(f+g) \wedge h \leqslant (f \wedge h) + (g \wedge h) \qquad (2)$$

Ex. Prove: If $a, b, c \in \mathbb{R}_+$, then $|a \wedge c - b \wedge c| \leqslant |a - b| \wedge c$

Proof: Fix $a, b \in \mathbb{R}_+$ and $c \in \mathbb{R}_+$

By (2), we have

$$a \wedge c = (a - b + b) \wedge c \leqslant (|a-b| + b) \wedge c \leqslant |a-b| \wedge c + b \wedge c$$

Thus, $a \wedge c - b \wedge c \leqslant |a-b| \wedge c$

Reversing the roles of a and b gives $b \wedge c - a \wedge c \leqslant |a-b| \wedge c$

Also, if $f, g, h \in \mathbb{R}_+^{\mathsf{X}}$, then

$$(f+g) \wedge h \leqslant (f \wedge h) + (g \wedge h) \tag{2}$$

Ex. Prove: If $a, b, c \in \mathbb{R}_+$, then $|a \wedge c - b \wedge c| \leqslant |a - b| \wedge c$

Proof: Fix $a, b \in \mathbb{R}_+$ and $c \in \mathbb{R}_+$

By (2), we have

$$a \wedge c = (a - b + b) \wedge c \leqslant (|a - b| + b) \wedge c \leqslant |a - b| \wedge c + b \wedge c$$

Thus, $a \wedge c - b \wedge c \leqslant |a - b| \wedge c$

Reversing the roles of a and b gives $b \wedge c - a \wedge c \leqslant |a - b| \wedge c$

Least and Greatest Elements

In dynamic programming, our aim is to maximize lifetime value

This is a function over the state space — not a number

Thus, the objective takes values in a partially ordered set (a set of functions over the state space)

This leads us to consider "least" and "greatest" elements, rather than traditional maxima and minima

Definitions follow...

Let $(P, \preceq)$ be a partially ordered set

Given $A \subset P$, we say that

- $\ell \in P$ is a **least element** of A if

$$\ell \in A \text{ and } a \succeq \ell \text{ for all } a \in A, \text{ and}$$

- $g \in P$ is a **greatest element** of A if

$$g \in A \text{ and } a \preceq g \text{ for all } a \in A.$$

Ex. Let P be any partially ordered set and fix $A \subset P$. Prove that A has at most one greatest element and at most one least element.

Example. The pointwise case

Let X be a nonempty set and V be a subset of $\mathbb{R}^{\mathsf{X}}$

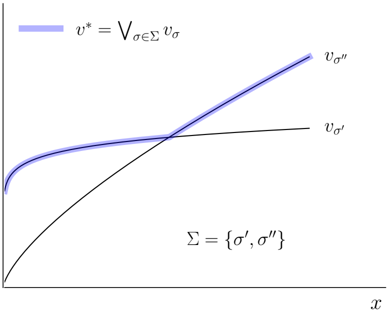
- V as a partially ordered set under the pointwise order $\leqslant$

Let $\{v_\sigma\} := \{v_\sigma\}_{\sigma \in \Sigma}$ be a finite collection of functions in V

Let $\vee_\sigma v_\sigma$ be the pointwise maximum

If $\vee_\sigma v_\sigma \in \{v_\sigma\}$, then $\vee_\sigma v_\sigma$ is the greatest element of $\{v_\sigma\}$

Example. The set $\{v_\sigma\}$ has no greatest element: neither $v_{\sigma'} \leqslant v_{\sigma''}$ nor $v_{\sigma''} \leqslant v_{\sigma'}$



Suprema and Infima in Partially Ordered Sets

Fix a partially ordered set $(P, \preceq)$ and nonempty subset A

In this setting,

- $u \in P$ is called an **upper bound** of A if $a \preceq u$ for all a in A
- $U_P(A) :=$ the set of all upper bounds of A in P

We call $\bar{u} \in P$ a **supremum** of A if $\bar{u}$ is a least element of $U_P(A)$

In other words,

$$\bar{u} \in U_P(A) \quad \text{and} \quad \bar{u} \preceq u \quad \text{for all} \quad u \in U_P(A)$$

Ex. Prove: A has at most one supremum in P

Proof: Suppose that s and s' are both suprema of A in P

Then both s and s' are upper bounds, so $s \preceq s'$ and $s' \preceq s$

Hence $s = s'$

Letting A be a subset of partially ordered space P,

- the supremum of A is typically denoted by $\bigvee A$
- If $A = \{a_i\}_{i \in I}$ we also write $\bigvee A$ as $\bigvee_i a_i$
- If $A = \{a, b\}$, then $\bigvee A$ is also written as $a \vee b$

Ex. Prove: A has at most one supremum in P

<u>Proof</u>: Suppose that s and s' are both suprema of A in P

Then both s and s' are upper bounds, so $s \preceq s'$ and $s' \preceq s$

Hence $s = s'$

Letting A be a subset of partially ordered space P,

- the supremum of A is typically denoted by $\bigvee A$
- If $A = \{a_i\}_{i \in I}$ we also write $\bigvee A$ as $\bigvee_i a_i$
- If $A = \{a, b\}$, then $\bigvee A$ is also written as $a \vee b$

We call $\ell \in P$ a **lower bound** of A if $\ell \preceq a$ for all a in A

An element $\bar{\ell}$ of P is called a **infimum** of A if

1. $\bar{\ell}$ is a lower bound of A and
2. $\ell \preceq \bar{\ell}$ for every lower bound ℓ of A

We use analogous notation to denote the infimum

For example, if $A = \{a, b\}$, then $\bigwedge A$ is also written as $a \wedge b$

Example. Let M be a nonempty set and let $\wp(M) =$ all subsets of M, partially ordered by $\subset$

Consider $\{A_i\}_{i\in I} \subset \wp(M)$

Ex. Prove that $\bigvee_i A_i = \cup_i A_i$ and $\bigwedge_i A_i = \cap_i A_i$

Proof: Observe that $A_j \subset \cup_i A_i$ for all $j \in I$

Hence $\cup_i A_i$ is an upper bound of $\{A_i\}$

Moreover, if $B \subset M$ and $A_j \subset B$ for all $i \in I$, then $\cup_i A_i \subset B$

This proves that $\cup_i A_i$ is the supremum

The proof of the infimum case is similar

Example. Let M be a nonempty set and let $\wp(M) =$ all subsets of M, partially ordered by $\subset$

Consider $\{A_i\}_{i\in I} \subset \wp(M)$

Ex. Prove that $\bigvee_i A_i = \cup_i A_i$ and $\bigwedge_i A_i = \cap_i A_i$

Proof: Observe that $A_j \subset \cup_i A_i$ for all $j \in I$

Hence $\cup_i A_i$ is an upper bound of $\{A_i\}$

Moreover, if $B \subset M$ and $A_j \subset B$ for all $i \in I$, then $\cup_i A_i \subset B$

This proves that $\cup_i A_i$ is the supremum

The proof of the infimum case is similar

Order-preserving maps

Let

- $(P, \preceq)$ and $(Q, \trianglelefteq)$ be partially ordered sets
- $T \colon P \to Q$

T is called **order-preserving** if, for all $x, y \in P$,

$$x \preceq y \quad \implies \quad Tx \trianglelefteq Ty$$

- Meaning: If x goes up then Tx goes up
- Very important concept for dynamic programming

Example. Let $(P, \preceq) = (\mathcal{C}, \leqslant)$ where

- $\mathcal{C}$ is all continuous functions from $[a, b]$ to $\mathbb{R}$
- $\leqslant$ is the pointwise order

If $I\colon \mathcal{C} \to \mathbb{R}$ is defined by

$$Ig := \int_a^b g(x)dx \qquad (g \in \mathcal{C})$$

then I is order-preserving on $\mathcal{C}$

(Larger functions have larger integrals)

Example. Let $\leqslant$ denote the pointwise order on $\mathbb{R}^n$

Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be defined by $Tx = Ax + b$

If $A \geqslant 0$, then T is order preserving on $\mathbb{R}^n$

Proof: Fix $x \leqslant y$

Then $0 \leqslant y - x$

$$\therefore \quad 0 \leqslant A(y - x) = Ay - Ax$$

$$\therefore \quad Ax \leqslant Ay$$

$$\therefore \quad Tx \leqslant Ty$$

Example. Let $\leqslant$ denote the pointwise order on $\mathbb{R}^n$

Let $T \colon \mathbb{R}^n \to \mathbb{R}^n$ be defined by $Tx = Ax + b$

If $A \geqslant 0$, then T is order preserving on $\mathbb{R}^n$

<u>Proof:</u> Fix $x \leqslant y$

Then $0 \leqslant y - x$

$$\therefore \quad 0 \leqslant A(y - x) = Ay - Ax$$

$$\therefore \quad Ax \leqslant Ay$$

$$\therefore \quad Tx \leqslant Ty$$

Let P, Q be partially ordered sets, $F\colon P \to Q$ be order-preserving

Suppose that $\{u_i\} \subset P$ has a greatest element

Ex. Prove that $\bigvee_i F u_i$ exists in Q and, moreover,

$$F \bigvee_i u_i = \bigvee_i F u_i$$

Proof: Let $\bar{u}$ be the greatest element of $\{u_i\}$

Then $F u_i \preceq F \bar{u}$ for all i

Hence $F \bar{u}$ is the greatest element and supremum of $\{F u_i\}$

That is, $\bigvee_i F u_i = F \bar{u} = F \bigvee_i u_i$

Let P, Q be partially ordered sets, $F\colon P \to Q$ be order-preserving

Suppose that $\{u_i\} \subset P$ has a greatest element

Ex. Prove that $\bigvee_i Fu_i$ exists in Q and, moreover,

$$F \bigvee_i u_i = \bigvee_i Fu_i$$

Proof: Let $\bar{u}$ be the greatest element of $\{u_i\}$

Then $Fu_i \preceq F\bar{u}$ for all i

Hence $F\bar{u}$ is the greatest element and supremum of $\{Fu_i\}$

That is, $\bigvee_i Fu_i = F\bar{u} = F \bigvee_i u_i$

Special Case: Real-Valued Functions

Special case: maps from $(P, \preceq)$ into $(\mathbb{R}, \leqslant)$

Then "order-preserving" = "increasing"

In particular, we also call $h \in \mathbb{R}^P$

- **increasing** if $x \preceq y$ implies $h(x) \leqslant h(y)$ and
- **decreasing** if $x \preceq y$ implies $h(x) \geqslant h(y)$

Let P be partially ordered by $\preceq$

We write $i\mathbb{R}^P$ **for the increasing functions in** $\mathbb{R}^P$

Thus,

$$h \in i\mathbb{R}^P \quad \iff \quad x, y \in P \text{ and } x \preceq y \text{ implies } h(x) \leqslant h(y)$$

Example. Let $P = \{1, \ldots, n\}$ and let $\preceq$ be the usual order $\leqslant$ on $\mathbb{R}$

Then

- $x \mapsto 2x$ and $x \mapsto \mathbb{1}\{2 \leqslant x\}$ are in $i\mathbb{R}^P$
- $x \mapsto -x$ and $x \mapsto \mathbb{1}\{x \leqslant 2\}$ are not

Ex. Prove the following:

If $f, g \in i\mathbb{R}^P$, then

- $\alpha f + \beta g \in i\mathbb{R}^P$ when $\alpha, \beta \geqslant 0$
- $f \vee g \in i\mathbb{R}^P$
- $f \wedge g \in i\mathbb{R}^P$

Ex. Given finite P, show that $i\mathbb{R}^P$ is closed in $\mathbb{R}^P$

Proof: Take $(f_k)_{k \geqslant 1}$ in $i\mathbb{R}^P$ and $f \in \mathbb{R}^P$ with $f_k \to f$

Since $f_k \to f$ we have $f_k(z) \to f(z)$ for all $z \in P$

- norm convergence implies pointwise convergence

Fix $x, y \in P$ with $x \preceq y$

From $(f_k) \subset i\mathbb{R}^P$ we have $f_k(x) \leqslant f_k(y)$ for all k

Since weak inequalities are preserved under limits, $f(x) \leqslant f(y)$

Hence $f \in i\mathbb{R}^P$

Strict inequalities

We write

- $f \ll g$ if $f(x) < g(x)$ for all $x \in$ some given set M
- $x \ll y$ if $x_i < y_i$ for all $i \in [n]$
- $A \ll B$ if $a_{ij} < b_{ij}$ for all i, j

These are not partial orders

Ex. Why is $f \ll g$ not a partial order on $\mathbb{R}^M$?

Order Isomorphisms

Let

- P and $\hat{P}$ be two partially ordered sets
- Φ be a map from P to $\hat{P}$

The map Φ is called

- an **order isomorphism** if Φ is bijective and Φ and Φ^{-1} are order-preserving, and
- an **order anti-isomorphism** if Φ is bijective and Φ and Φ^{-1} are order-reversing.

Example. If $P = \hat{P} = \mathbb{R}^n_+$, then $p \mapsto p^2$ is an order isomorphism

A partially ordered set $V = (V, \preceq)$ is called a **lattice** if

$$u, v \in V \implies u \vee v \in V \text{ and } u \wedge v \in V$$

- by induction, the sup and inf of any finite set also exist in V

A subset S of a lattice V is called a **sublattice** of V if

$$u, v \in S \implies u \vee v \in S \text{ and } u \wedge v \in S$$

Example. Let $C[a, b] :=$ all continuous $f\colon [a, b] \to \mathbb{R}$

$C[a, b]$ is a sublattice of $(\mathbb{R}^{[a,b]}, \leqslant)$, since given $f, g \in C[a, b]$, the functions $f \vee g$ and $f \wedge g$ are also continuous

Let L and $\hat{L}$ be two lattices and let $\{v_i\}_{i \in I}$ be a finite subset of L

Ex. Prove the following statements:

1. If F is an order isomorphism from L to $\hat{L}$, then

$$F \bigvee_i v_i = \bigvee_i F v_i \quad \text{and} \quad F \bigwedge_i v_i = \bigwedge_i F v_i$$

2. If F is an order anti-isomorphism from L to $\hat{L}$, then

$$F \bigvee_i v_i = \bigwedge_i F v_i \quad \text{and} \quad F \bigwedge_i v_i = \bigvee_i F v_i$$

Blackwell's Condition

Fix $U \subset \mathbb{R}^{\mathsf{X}}$ with X finite

Assume $u \in U$ and $c \in \mathbb{R}_+$ implies $u + c \in U$

Let T be an order preserving self-map on U

Lemma. If there exists a constant $\beta \in (0, 1)$ such that

$$T(u + c) \leqslant Tu + \beta c \quad \text{for all } u \in U \text{ and } c \in \mathbb{R}_+$$

then T is a contraction of modulus β on U w.r.t. $\| \cdot \|_\infty$

Proof:

Let U, T have the stated properties and fix $u, v \in U$

We have

$$\begin{aligned} Tu &= T(v + u - v) \\ &\leqslant T(v + \|u - v\|_\infty) \\ &\leqslant Tv + \beta\|u - v\|_\infty \end{aligned}$$

Hence

$$Tu - Tv \leqslant \beta\|u - v\|_\infty$$

Reversing the roles of u and v proves the claim

(First Order) Stochastic Dominance

Some dynamic programs have useful monotonicity properties

And some of these results use an order over distributions

Examples.

- One wage offer distribution is "better" than another
- One investment opportunity is "better" than another

The most important of these partial orders is called "first order stochastic dominance"

In this section we define it

To start, let's consider ordering distributions in a special case

Example. $X \sim B(n, 0.5)$ counts heads in n flips of a fair coin

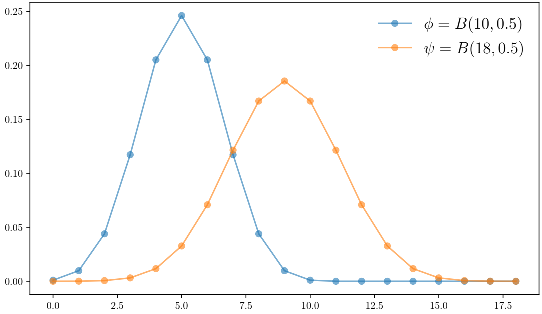
Suppose

- $\varphi \stackrel{d}{=} X \sim B(10, 0.5)$ and
- $\psi \stackrel{d}{=} Y \sim B(18, 0.5)$

- Y counts over more flips, so it should be "larger"

Hence we expect φ is "$\preceq$" ψ in some sense

Distribution ψ seems “larger than” φ — more mass on higher draws



But how can we make this idea precise?

Let X be a finite set partially ordered by $\preceq$

Fix $\varphi, \psi \in \mathcal{D}(\mathsf{X})$

Write $\langle u, \varphi \rangle$ for $\sum_x u(x)\varphi(x)$, etc.

We say that ψ **stochastically dominates** φ and write $\varphi \preceq_{\mathbf{F}} \psi$ if

$$u \in i\mathbb{R}^{\mathsf{X}} \quad \implies \quad \langle u, \varphi \rangle \leqslant \langle u, \psi \rangle$$

Example. If

- $\varphi \stackrel{d}{=} X \sim B(10, 0.5)$ and
- $\psi \stackrel{d}{=} Y \sim B(18, 0.5)$,

then $\varphi \preceq_{\mathrm{F}} \psi$

Proof: Fix $u \in i\mathbb{R}^{\mathsf{X}}$ and let

- $\mathsf{X} = \{0, \ldots, 18\}$ and
- $W_1, \ldots, W_{18}$ be IID Bernoulli with $\mathbb{P}\{W_i = 1\} = 0.5$ for all i

Then $X := \sum_{i=1}^{10} W_i \stackrel{d}{=} \varphi$ and $Y := \sum_{i=1}^{18} W_i \stackrel{d}{=} \psi$

Clearly $X \leqslant Y$ with probability one (i.e., for any draw of $\{W_i\}_{i=1}^{18}$)

Hence $u(X) \leqslant u(Y)$

Hence $\mathbb{E} u(X) \leqslant \mathbb{E} u(Y)$

In other words,

$$\langle u, \varphi \rangle \leqslant \langle u, \psi \rangle$$

Example. An agent has preferences over outcomes in X

Preferences are determined by a utility function $u \in \mathbb{R}^{\mathsf{X}}$

The agent prefers more to less, so $u \in i\mathbb{R}^{\mathsf{X}}$

Suppose that the agent ranks lotteries over X according to expected utility

- evaluates $\varphi \in \mathcal{D}(\mathsf{X})$ according to $\sum_x u(x)\varphi(x)$

Then the agent (weakly) prefers ψ to φ whenever $\varphi \preceq_{\mathrm{F}} \psi$

Another Perspective

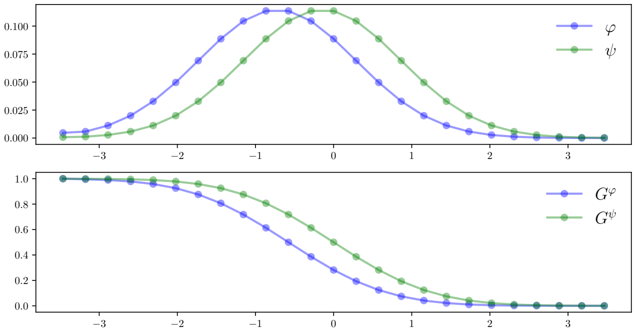
Given $\varphi \in \mathcal{D}(\mathsf{X})$, let

$$G^{\varphi}(y) := \sum_{x \in \mathsf{X}} \mathbb{1}\{y \preceq x\}\varphi(x) \qquad (y \in \mathsf{X})$$

This is the **counter** CDF of φ

Lemma. For each $\varphi, \psi \in \mathcal{D}(\mathsf{X})$, the following statements hold:

1. $\varphi \preceq_{\mathrm{F}} \psi \implies G^{\varphi} \leqslant G^{\psi}$
2. If X is totally ordered by $\preceq$, then $G^{\varphi} \leqslant G^{\psi} \implies \varphi \preceq_{\mathrm{F}} \psi$



0.100
0.075
0.050
0.025
0.000
−3
−2
−1
0
1
2
3
φ
ψ
1.0
0.8
0.6
0.4
0.2
0.0
G^{φ}
G^{ψ}

Lemma. $\preceq_{\mathrm{F}}$ is a partial order on $\mathcal{D}(\mathsf{X})$

Proof:

Let's just prove transitivity

Suppose $f, g, h \in \mathcal{D}(\mathsf{X})$ with $f \preceq_{\mathrm{F}} g$ and $g \preceq_{\mathrm{F}} h$

Fixing $u \in i\mathbb{R}^{\mathsf{X}}$, we have

$$\langle u, f \rangle \leqslant \langle u, g \rangle \quad \text{and} \quad \langle u, g \rangle \leqslant \langle u, h \rangle$$

Hence $\langle u, f \rangle \leqslant \langle u, h \rangle$

Since u was arbitrary in $i\mathbb{R}^{\mathsf{X}}$, we are done

Parametric Monotonicity

Let $(P, \preceq)$ be a partially ordered set

Given two self-maps S and T on P, we set

$$S \preceq T \iff Sx \preceq Tx \text{ for every } x \in P$$

We say that T **dominates** S on P

Ex. Show that $\preceq$ is a partial order on

$$\mathcal{S}_P := P^P := \text{ set of all self-maps on } P$$

Proof of antisymmetry of $\preceq$ on $\mathcal{S}_P$:

Let $(P, \preceq)$ and $S, T \in \mathcal{S}_P$ be as defined above

Suppose $S \preceq T$ and $T \preceq S$

Fix any $x \in P$

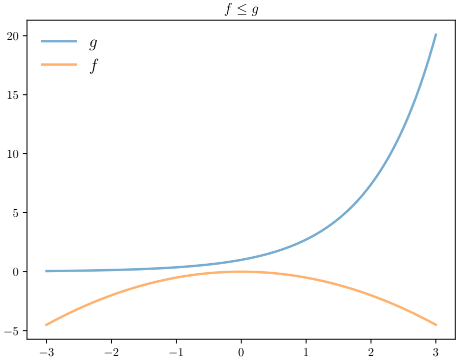
We have $Sx \preceq Tx$ and $Tx \preceq Sx$

Since $\preceq$ is antisymmetric on P, we have $Sx = Tx$

Since p was arbitrary, $S = T$

Hence $\preceq$ is antisymmetric on $\mathcal{S}_P$

Example. If $(\preceq, P) = (\leqslant, \mathbb{R})$, then $\leqslant$ is the pointwise order over functions



Example. Consider $\mathbb{R}^n_+$ with the pointwise order $\leqslant$

- Called the **positive cone** in $\mathbb{R}^n$

Let

- $Sx = Ax + b$
- $Tx = Bx + b$

Ex. Show that $0 \leqslant A \leqslant B \implies T$ dominates S on $\mathbb{R}^n_+$

Proof: Fixing $x \in \mathbb{R}^n_+$, suffices to show that $Sx \leqslant Tx$

Since $A \leqslant B$ and $x \geqslant 0$, we have $Ax \leqslant Bx$

Hence $Sx \leqslant Tx$

Example. Consider $\mathbb{R}^n_+$ with the pointwise order $\leqslant$

- Called the **positive cone** in $\mathbb{R}^n$

Let

- $Sx = Ax + b$
- $Tx = Bx + b$

Ex. Show that $0 \leqslant A \leqslant B \implies T$ dominates S on $\mathbb{R}^n_+$

Proof: Fixing $x \in \mathbb{R}^n_+$, suffices to show that $Sx \leqslant Tx$

Since $A \leqslant B$ and $x \geqslant 0$, we have $Ax \leqslant Bx$

Hence $Sx \leqslant Tx$

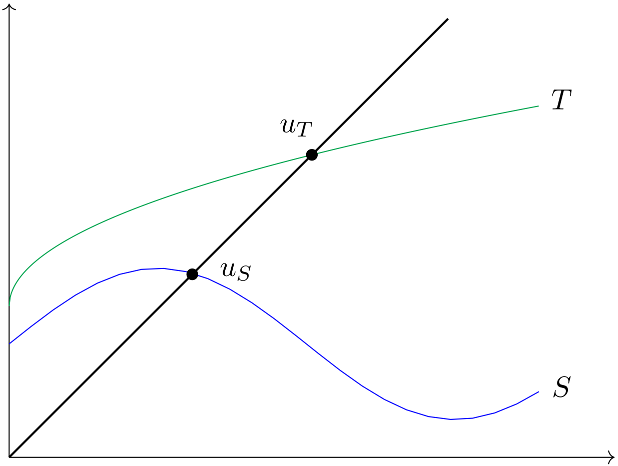
Conjecture: If $S \leqslant T$, then the fixed points of T will be larger

This is not true in general...

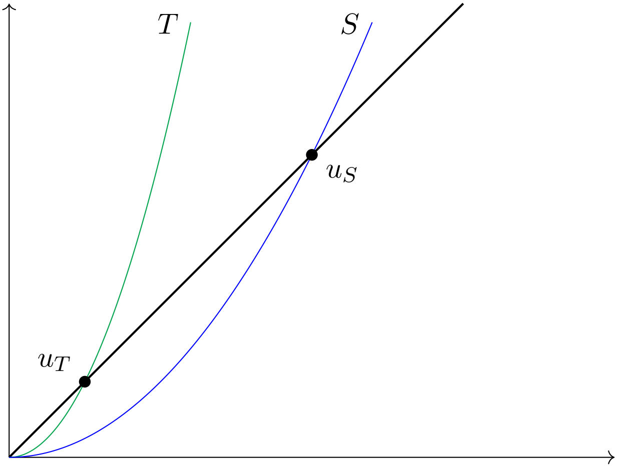
Conjecture: If $S \leqslant T$, then the fixed points of T will be larger

This is not true in general...

Sometimes true:



And sometimes false:



One difference: in the first case, T is globally stable

This leads us to our next result

Proposition. Let

- S and T be self-maps on $M \subset \mathbb{R}^n$
- $\leqslant$ be the pointwise order on M

If

1. T dominates S on M and
2. T is order-preserving and globally stable on M,

then the unique fixed point of T dominates any fixed point of S

Proof: Assume the conditions

Let

- u_T be the unique fixed point of T and
- u_S be any fixed point of S

Since $S \leqslant T$, we have $u_S = Su_S \leqslant Tu_S$

Applying T to both sides of $u_S \leqslant Tu_S$ gives

$$u_S \leqslant Tu_S \leqslant T^2 u_S$$

Continuing in this fashion yields $u_S \leqslant T^k u_S$ for all $k \in \mathbb{N}$

Since $\leqslant$ is preserved under limits and T is globally stable,

$$u_S \leqslant \lim_k T^k u_S = u_T$$

Example. Recall that, in the job search model,

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

We found h^* as the fixed point of $g\colon \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$g(h) = c + \beta \sum_{w'} \max\left\{\frac{w'}{1-\beta}, h\right\} \varphi(w')$$

In the exercise, you showed that g is a contraction map on $\mathbb{R}_+$

Ex. Prove that the optimal continuation value h^* is increasing in β

Proof: Fix $\beta_1 \leqslant \beta_2$ and let

- $h_i^* :=$ fixed point corresponding to β_i
- $g_i :=$ fixed point map corresponding to β_i

Since $\beta_1 \leqslant \beta_2$, we have $g_1(h) \leqslant g_2(h)$ for all $h \in \mathbb{R}_+$

In addition, g_2 is

1. a contraction (so globally stable) and
2. increasing (order-preserving)

Hence $h_1^* \leqslant h_2^*$

Ex. Prove that the optimal continuation value h^* is increasing in β

Proof: Fix $\beta_1 \leqslant \beta_2$ and let

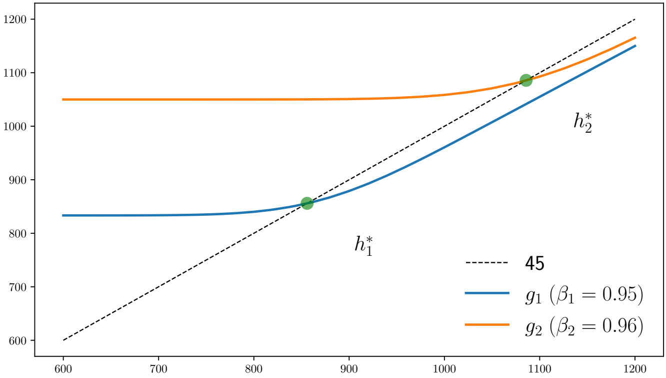
- $h_i^* :=$ fixed point corresponding to β_i
- $g_i :=$ fixed point map corresponding to β_i

Since $\beta_1 \leqslant \beta_2$, we have $g_1(h) \leqslant g_2(h)$ for all $h \in \mathbb{R}_+$

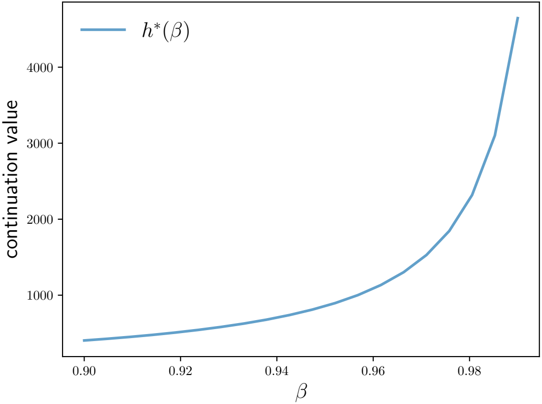
In addition, g_2 is

1. a contraction (so globally stable) and
2. increasing (order-preserving)

Hence $h_1^* \leqslant h_2^*$


1200
1100
1000
900
800
700
600
600
700
800
900
1000
1100
1200
h_1^*
h_2^*
45
g_1 ($\beta_1 = 0.95$)
g_2 ($\beta_2 = 0.96$)

Ex. Replicate this figure



Reminders

<u>Def</u>. $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ is an **eigenpair** of $n \times n$ matrix A if

$$v \neq 0 \quad \text{and} \quad Av = \lambda v$$

The **eigenspace** of eigenvalue λ is

$$E_\lambda := \{w \in \mathbb{C}^n : w = 0 \text{ or } (\lambda, w) \text{ is an eigenpair of } A\}$$

Ex. Show that E_λ is a linear subspace of $\mathbb{C}^n$

<u>Proof</u>: If $v, w \in E_\lambda$ and $\alpha, \beta \in \mathbb{C}$, then

$$A(\alpha w + \beta v) = \alpha A w + \beta A v = \alpha \lambda w + \beta \lambda v = \lambda(\alpha w + \beta v)$$

Reminders

<u>Def</u>. $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ is an **eigenpair** of $n \times n$ matrix A if

$$v \neq 0 \quad \text{and} \quad Av = \lambda v$$

The **eigenspace** of eigenvalue λ is

$$E_\lambda := \{w \in \mathbb{C}^n : w = 0 \text{ or } (\lambda, w) \text{ is an eigenpair of } A\}$$

Ex. Show that E_λ is a linear subspace of $\mathbb{C}^n$

<u>Proof</u>: If $v, w \in E_\lambda$ and $\alpha, \beta \in \mathbb{C}$, then

$$A(\alpha w + \beta v) = \alpha Aw + \beta Av = \alpha\lambda w + \beta\lambda v = \lambda(\alpha w + \beta v)$$

Implication: exists a continuum of eigenvectors paired with λ

So what can we say about uniqueness?

Let (λ, v) be an eigenpair for A

Def. v has (geometric) **multiplicity one** if $\dim E_\lambda = 1$

In other words,

$$w \in E_\lambda \implies w = \alpha v \text{ for some } \alpha \in \mathbb{C}$$

In a sense, there is "just one" eigenvector corresponding to λ, since any other is a scalar multiple

Nonnegative Matrices

Def. Matrix A is called

- **nonnegative** if $A \geqslant 0$
- **positive** if $A \gg 0$
- **irreducible** if it is square, nonnegative and

$$\sum_{k=1}^{\infty} A^k \gg 0$$

For square A,

$$\text{positive} \implies \text{irreducible} \implies \text{nonnegative}$$

Let A be square

It is <u>not</u> always true that $\rho(A)$ is an eigenvalue of A

Example. Let

$$A := \begin{pmatrix} -1 & 0 \\ 0 & 1/2 \end{pmatrix}$$

The **spectrum** (set of eigenvalues) of A is

$$\sigma(A) = \{-1, 1/2\}$$

Hence $\rho(A) = |-1| = 1 \notin \sigma(A)$

However, when $A \geqslant 0$, we have the following result

Theorem. (Perron–Frobenius) If $A \geqslant 0$, then $\rho(A)$ is an eigenvalue of A with nonnegative, real-valued right and left eigenvectors

In particular, there exists

- a nonnegative, nonzero column vector e s.t. $Ae = \rho(A)e$
- a nonnegative, nonzero row vector ε s.t. $\varepsilon A = \rho(A)\varepsilon$

If A **is irreducible**, then these eigenvectors are everywhere positive and have multiplicity of one

If A **is positive**, then with e and ε such that $\langle \varepsilon, e \rangle = 1$, we have

$$\rho(A)^{-t} A^t \to e\,\varepsilon \qquad (t \to \infty)$$

In this setting,

- $\rho(A)$ is also called the **dominant eigenvalue**
- e is called the **dominant right eigenvector**
- ε is called the **dominant left eigenvector**

Note also

$$\varepsilon A = \rho(A)\varepsilon \quad \iff \quad A^\top \varepsilon^\top = \rho(A)\varepsilon^\top$$

Hence $\varepsilon^\top$ is the dominant right eigenvector of $A^\top$

Since the dominant eigenvectors are only defined up to constant multiples, we often normalize so that $\langle \varepsilon, e \rangle = 1$

Let's check these results for arbitrary positive A

```
julia> using LinearAlgebra

julia> A = [0.3 0.9;
            1.0 0.1];

julia> λ_1, λ_2 = eigvals(A)
2-element Vector{Float64}:
 -0.7539392014169456
  1.1539392014169458

julia> rA = λ_2    # r(A) is the positive eigenvalue
1.1539392014169458
```

```
julia> right_evecs = eigvecs(A)
2×2 Matrix{Float64}:
 -0.649386  0.725426
  0.760459  0.6883

julia> e = right_evecs[:, 2]   # dominant right eigenvector
2-element Vector{Float64}:
 0.7254262498099013
 0.6882999027217298

julia> left_evs = eigvecs(A')  # transpose to get left eigenvector
2×2 Matrix{Float64}:
 -0.6883    0.760459
  0.725426  0.649386

julia> ϵ = left_evs[:, 2]'     # dominant left eigenvector
1×2 adjoint(::Vector{Float64}) with eltype Float64:
 0.760459  0.649386
```

```
#  Checking the eigenpair relations

 julia> A * e
2-element Vector{Float64}:
 0.8370977873925273
 0.7942562400820743

julia> rA * e
2-element Vector{Float64}:
 0.8370977873925274
 0.7942562400820744


julia> ϵ * A
1×2 adjoint(::Vector{Float64}) with eltype Float64:
 0.877524  0.749352

julia> rA * ϵ
1×2 adjoint(::Vector{Float64}) with eltype Float64:
 0.877524  0.749352
```

```
#  The matrix A is everywhere positive
#
#  Hence we expect, for large k,
#
#       r(A)^(-k) * A^k ≈ e ϵ


julia> k = 1000
1000

julia> rA^(-k) * A^k
2×2 Matrix{Float64}:
 0.552414  0.471728
 0.524142  0.447586

julia> e * ϵ
2×2 Matrix{Float64}:
 0.551657  0.471082
 0.523424  0.446972
```

Bounds on the spectral radius

Fix $n \times n$ matrix A and set

- $\mathrm{rs}_i(A) :=$ the i-th row sum of A and
- $\mathrm{cs}_j(A) :=$ the j-th column sum of A

Corollary. If $A \geqslant 0$, then

1. $\min_i \mathrm{rs}_i(A) \leqslant \rho(A) \leqslant \max_i \mathrm{rs}_i(A)$ and
2. $\min_j \mathrm{cs}_j(A) \leqslant \rho(A) \leqslant \max_j \mathrm{cs}_j(A)$

Ex. Prove this via the PF theorem

Proof for the column sum case

Fix $A \geqslant 0$ and let e be the dominant right eigenvector

We normalize e by setting $\mathbb{1}^\top e = \sum_j e_j = 1$

From $\rho(A)e = Ae$ we have

$$\rho(A) = \rho(A)\mathbb{1}^\top e = \mathbb{1}^\top(\rho(A)e) = \mathbb{1}^\top Ae = \sum_j \mathrm{cs}_j(A)e_j$$

Therefore, $\rho(A)$ is a weighted average of the column sums

Hence $\min_j \mathrm{cs}_j(A) \leqslant \rho(A) \leqslant \max_j \mathrm{cs}_j(A)$

Stochastic Matrices

Let P be a square matrix

Def. P is called **stochastic** if $P \geqslant 0$ and $P\mathbb{1} = \mathbb{1}$

Ex. Show that P is stochastic $\implies \rho(P) = 1$

Row vector ψ is called a **stationary distribution** of P if

$$\psi \geqslant 0, \quad \psi\mathbb{1} = 1 \quad \text{and} \quad \psi P = \psi$$

Stationary distributions very important for Markov dynamics...

Existence of Stationary Distributions

Let P be a stochastic matrix

Ex. Prove: P has at least one stationary distribution

Proof: By the PF theorem,

$\exists$ a nonzero, nonnegative row vector φ satisfying $\varphi P = \varphi$

Since φ is nonzero, $\varphi \mathbb{1} > 0$

Setting $\psi := \varphi / (\varphi \mathbb{1})$ gives the desired vector

Existence of Stationary Distributions

Let P be a stochastic matrix

Ex. Prove: P has at least one stationary distribution

Proof: By the PF theorem,

$$\exists \text{ a nonzero, nonnegative row vector } \varphi \text{ satisfying } \varphi P = \varphi$$

Since φ is nonzero, $\varphi \mathbb{1} > 0$

Setting $\psi := \varphi / (\varphi \mathbb{1})$ gives the desired vector

Uniqueness of Stationary Distributions

Ex. Prove: If P is also irreducible, then the stationary vector ψ is everywhere positive and unique

Proof of Positivity: See Perron–Frobenius theorem

Proof of Uniqueness: Let $\varphi \geqslant 0$ satisfy $\varphi \mathbb{1} = 1$ and $\varphi P = \varphi$

By the Perron–Frobenius theorem, $\varphi = \alpha \psi$ for some $\alpha > 0$

But then $1 = \varphi \mathbb{1} = \alpha \psi \mathbb{1} = \alpha$

Hence $\varphi = \psi$

Uniqueness of Stationary Distributions

Ex. Prove: If P is also irreducible, then the stationary vector ψ is everywhere positive and unique

Proof of Positivity: See Perron–Frobenius theorem

Proof of Uniqueness: Let $\varphi \geqslant 0$ satisfy $\varphi \mathbb{1} = 1$ and $\varphi P = \varphi$

By the Perron–Frobenius theorem, $\varphi = \alpha \psi$ for some $\alpha > 0$

But then $1 = \varphi \mathbb{1} = \alpha \psi \mathbb{1} = \alpha$

Hence $\varphi = \psi$

Uniqueness of Stationary Distributions

Ex. Prove: If P is also irreducible, then the stationary vector ψ is everywhere positive and unique

Proof of Positivity: See Perron–Frobenius theorem

Proof of Uniqueness: Let $\varphi \geqslant 0$ satisfy $\varphi \mathbb{1} = 1$ and $\varphi P = \varphi$

By the Perron–Frobenius theorem, $\varphi = \alpha \psi$ for some $\alpha > 0$

But then $1 = \varphi \mathbb{1} = \alpha \psi \mathbb{1} = \alpha$

Hence $\varphi = \psi$

```
julia> P = [0.2 0.8;
            0.1 0.9]
2×2 Matrix{Float64}:
    0.2  0.8
    0.1  0.9

julia> using QuantEcon

julia> mc = MarkovChain(P)
    Discrete Markov Chain
    stochastic matrix of type Matrix{Float64}:
    [0.2 0.8; 0.1 0.9]

julia> is_irreducible(mc)
true

julia> stationary_distributions(mc)
    1-element Vector{Vector{Float64}}:
    [0.1111111111111111, 0.8888888888888888]
```

Lake Model of Employment

An illustration of the Perron–Frobenius theorem

We analyze a model of employment and unemployment flows in a large population

The model is sometimes called a "lake model"

Two "pools" of workers:

- those who are currently employed and
- those who are currently unemployed but still seeking work

FP theorem helps us analyze dynamics

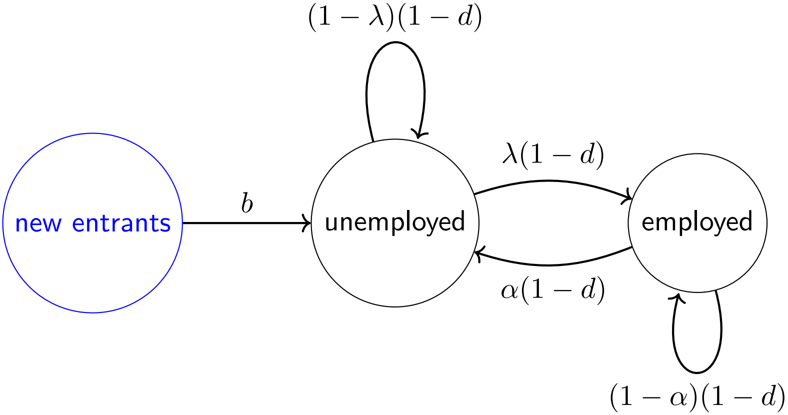
Workers

- **exit** the workforce at rate d
- **enter** the workforce at rate b
- **separate** from their jobs at rate α
- **find jobs** at rate λ

Assumptions:

- All parameters lie in $(0, 1)$
- New workers are initially unemployed

Transition rates:



Let

- $u_t :=$ number of **unemployed workers** at time t
- $e_t :=$ number of **employed workers**
- $n_t := e_t + u_t :=$ total **population** of workers

Dynamics are

$$u_{t+1} = (1-d)\alpha e_t + (1-d)(1-\lambda)u_t + bn_t$$

$$e_{t+1} = (1-d)(1-\alpha)e_t + (1-d)\lambda u_t$$

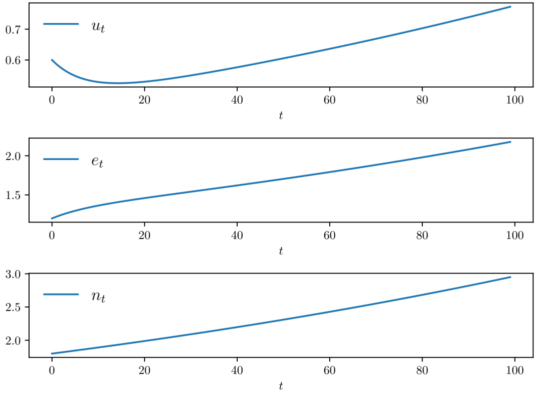


Figure: Example simulation when $b > d$ (population growth)

Can we say more about the dynamics of this system?

For example,

- what long run unemployment rate should we expect?
- do outcomes depend on the initial conditions u_0 and e_0?
- Or are there general statements we can make?

We define

$$x_t := \begin{pmatrix} u_t \\ e_t \end{pmatrix}$$

and

$$A := \begin{pmatrix} (1-d)(1-\lambda)+b & (1-d)\alpha+b \\ (1-d)\lambda & (1-d)(1-\alpha) \end{pmatrix}$$

Dynamics can now be written

$$x_{t+1} = Ax_t$$

Hence

$$x_t = A^t x_0 \quad \text{where} \quad x_0 = \begin{pmatrix} u_0 \\ e_0 \end{pmatrix}$$

Ex. With $g := b - d$, show that $n_{t+1} = (1+g) n_t$ for all t

Proof: The column sums of A are

$$(1-d)(1-\lambda) + b + (1-d)\lambda = 1 + g$$

and

$$(1-d)\alpha + b + (1-d)(1-\alpha) = 1 + g$$

From $x_{t+1} = A x_t$ and $n_t = u_t + e_t$ we have

$$n_{t+1} = \mathbb{1}^\top x_{t+1} = \mathbb{1}^\top A x_t = (1+g) \mathbb{1}^\top x_t = (1+g) n_t$$

Ex. With $g := b - d$, show that $n_{t+1} = (1+g)n_t$ for all t

<u>Proof</u>: The column sums of A are

$$(1-d)(1-\lambda) + b + (1-d)\lambda = 1 + g$$

and

$$(1-d)\alpha + b + (1-d)(1-\alpha) = 1 + g$$

From $x_{t+1} = Ax_t$ and $n_t = u_t + e_t$ we have

$$n_{t+1} = \mathbb{1}^\top x_{t+1} = \mathbb{1}^\top A x_t = (1+g)\mathbb{1}^\top x_t = (1+g)n_t$$

Ex. Prove that $\rho(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant \rho(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant \rho(A) \leqslant 1 + g$

PF theorem $\implies$ $1 + g$ is the dominant eigenvalue of A

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = \begin{pmatrix} 1+g & 1+g \end{pmatrix} = \rho(A)\mathbb{1}^\top$$

Ex. Prove that $\rho(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant \rho(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant \rho(A) \leqslant 1 + g$

PF theorem $\implies$ $1 + g$ is the dominant eigenvalue of A

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = \begin{pmatrix} 1+g & 1+g \end{pmatrix} = \rho(A)\mathbb{1}^\top$$

Ex. Prove that $\rho(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant \rho(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant \rho(A) \leqslant 1 + g$

PF theorem $\implies$ $1 + g$ is the dominant eigenvalue of A

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = \begin{pmatrix} 1+g & 1+g \end{pmatrix} = \rho(A)\mathbb{1}^\top$$

Ex. Prove that $\rho(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant \rho(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant \rho(A) \leqslant 1 + g$

PF theorem $\implies$ $1 + g$ is the dominant eigenvalue of A

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = \begin{pmatrix} 1+g & 1+g \end{pmatrix} = \rho(A)\mathbb{1}^\top$$

Ex. Prove that

$$\bar{x} := \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix}$$

with

$$\bar{u} := \frac{1+g-(1-d)(1-\alpha)}{1+g-(1-d)(1-\alpha)+(1-d)\lambda} \quad \text{and} \quad \bar{e} := 1-\bar{u}$$

is the dominant right eigenvector of A

Proof: Just show $A\bar{x} = (1+g)\bar{x}$

Ex. Prove that

$$\bar{x} := \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix}$$

with

$$\bar{u} := \frac{1+g-(1-d)(1-\alpha)}{1+g-(1-d)(1-\alpha)+(1-d)\lambda} \quad \text{and} \quad \bar{e} := 1-\bar{u}$$

is the dominant right eigenvector of A

Proof: Just show $A\bar{x} = (1+g)\bar{x}$

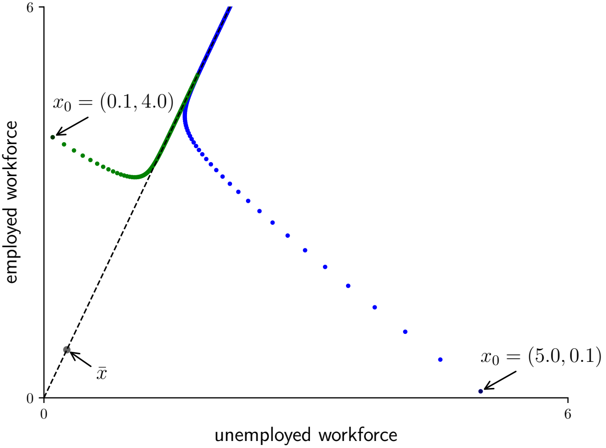
```julia
using LinearAlgebra

α, λ, d, b = 0.01, 0.1, 0.02, 0.025
g = b - d
A = [(1 - d) * (1 - λ) + b     (1 - d) * α + b;
     (1 - d) * λ               (1 - d) * (1 - α)]

ū = (1 + g - (1 - d) * (1 - α)) /
        (1 + g - (1 - d) * (1 - α) + (1 - d) * λ)

ē = 1 - ū
x̄ = [ū; ē]

println(isapprox(A * x̄, (1 + g) * x̄))   # prints true
```


6
0
0
6
employed workforce
unemployed workforce
$x_0 = (0.1, 4.0)$
$x_0 = (5.0, 0.1)$
$\bar{x}$

Let

$$D := \{x \in \mathbb{R}^2 : x = \alpha \bar{x} \text{ for some } \alpha > 0\}$$

- Shown as a dashed black line in the last figure
- The two time paths are of the form $(x_t)_{t \geqslant 0} = (A^t x_0)_{t \geqslant 0}$
- In both cases, the paths converge to D over time

Suggests all paths are "eventually almost" multiples of $\bar{x}$

How can we explain this strong regularity?

From the Perron–Frobenius theorem, since $A \gg 0$, we have

$$A^t \approx \rho(A)^t \cdot \bar{x}\mathbb{1}^\top = (1+g)^t \begin{pmatrix} \bar{u} & \bar{u} \\ \bar{e} & \bar{e} \end{pmatrix} \quad \text{for large } t$$

Hence, $\forall\, x_0 = (u_0\ e_0)^\top$,

$$A^t x_0 \approx (1+g)^t \begin{pmatrix} \bar{u} & \bar{u} \\ \bar{e} & \bar{e} \end{pmatrix} \begin{pmatrix} u_0 \\ e_0 \end{pmatrix}$$

$$= (1+g)^t (u_0 + e_0) \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix} = n_t \bar{x},$$

where $n_t = (1+g)^t n_0$ and $n_0 = u_0 + e_0$

Regardless of x_0, state scales along $\bar{x}$ at rate of population growth

Rates

Unemployment rate $= u_t / n_t$

For large t, we have $u_t \approx n_t \bar{u}$

Hence unemployment rate $\approx (n_t \bar{u}) / n_t = \bar{u}$

Hence $\bar{u}$ <u>is the long run rate of unemployment</u>

Similarly, $\bar{e}$ is the long run employment rate

$\implies$ dominant eigenvector gives unemployment rates

Extensions

Further analysis: how are α, λ, b and d determined?

For the hiring rate λ, we could use the job search model

In particular, with w^* as the reservation wage, we could set

$$\lambda = \mathbb{P}\{w_t \geqslant w^*\} = \sum_{w \in \mathsf{W}} \varphi(w) \mathbb{1}\{w \geqslant w^*\}$$

Doing so would allow us to study the crucial rate λ in terms of fundamental primitives, such as

- unemployment compensation
- impatience of individual agents, etc.

Linear Operators

An $n \times n$ matrix $A = (a_{ij})$ is

1. an $n \times n$ array of (real) numbers a_{ij}
2. a linear operator from $\mathbb{R}^k$ to $\mathbb{R}^n$ mapping $u \mapsto Au$

The matrix representation is important

But the linear operator representation is more fundamental

Let's clarify these ideas

A **linear operator** on $\mathbb{R}^n$ is a map L from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that

$$L(\alpha u + \beta v) = \alpha L u + \beta L v \quad \text{for all } u, v \in \mathbb{R}^n \text{ and } \alpha, \beta \in \mathbb{R}$$

Ex. Prove: if A is an $n \times n$ matrix, then $u \mapsto Au$ defines a linear operator

In fact the converse is also true:

Theorem. If L is a linear operator on $\mathbb{R}^n$, then there exists an $n \times n$ matrix $A = (a_{ij})$ such that $Lu = Au$ for all $u \in \mathbb{R}^n$

Proof: See, e.g., this link

Linear operators on function space

Let $\mathsf{X} = \{x_1, \dots, x_n\}$

A **linear operator** on $\mathbb{R}^{\mathsf{X}}$ is a self-map L on $\mathbb{R}^{\mathsf{X}}$ such that, for all $u, v \in \mathbb{R}^{\mathsf{X}}$ and $\alpha, \beta \in \mathbb{R}$, we have

$$L(\alpha u + \beta v) = \alpha L u + \beta L v$$

- acts as a linear operator on $\mathbb{R}^n$ when elements of $\mathbb{R}^{\mathsf{X}}$ are viewed as vectors

Below,

$$\mathcal{L}(\mathbb{R}^{\mathsf{X}}) := \text{all linear operators on } \mathbb{R}^{\mathsf{X}}$$

Ex. Let ℓ be a map from $\mathsf{X} \times \mathsf{X}$ to $\mathbb{R}$

Show that L defined by

$$(Lu)(x) = \sum_{x' \in \mathsf{X}} \ell(x, x') u(x') \qquad \left(u \in \mathbb{R}^{\mathsf{X}}\right)$$

is in $\mathcal{L}(\mathbb{R}^{\mathsf{X}})$

In fact every $L \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ takes the form above

The proof is as follows:

1. The "kernel" ℓ is just a matrix
2. In finite dimensions, linear operator $\leftrightarrow$ matrix

Let's summarize (assuming $\mathsf{X} = \{x_1, \ldots, x_n\}$)

The following sets are in one-to-one correspondence

(a) The set of all $n \times n$ real matrices

(b) The set of all linear operators on $\mathbb{R}^n$

(c) The set $\mathcal{L}(\mathbb{R}^{\mathsf{X}})$ of linear operators on $\mathbb{R}^{\mathsf{X}}$

(d) The set of all functions from $\mathsf{X} \times \mathsf{X}$ to $\mathbb{R}$

That said, linear operators are more general

- extend to infinite dims
- extend to abstract vector space

Also, the matrix representation can be

1. tedious to construct and
2. difficult to instantiate in memory in large problems

Example. Suppose $\mathsf{X} = \mathsf{Y} \times \mathsf{Z}$ and let $n = |\mathsf{Y}| \times |\mathsf{Z}|$

Fix $Q \colon \mathsf{Z} \times \mathsf{Z} \to \mathbb{R}$ and consider

$$(Lu)(x) = (Lu)(y, z) = \sum_{z'} u(y, z') Q(z, z')$$

Ex. Show that $L \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$

Hence L can be represented as an $n \times n$ matrix

Choose between

1. **column-major** order (Julia, Fortran, Matlab, etc.)
2. **row-major** order (Python, C, etc.)

But

- an $n \times n$ matrix has to be instantiated in memory, even though the operation is only an inner product in Z
- construction is tedious / error-prone
- confusion when swapping between column- and row-major orderings

Fortunately, some modern scientific computing environments support linear operators directly

- defining linear operators
- providing linear algebra routines (inversion, etc.)

In what follows we take an operator-centric approach

Positive operators

Let $\mathbb{R}^X_+ :=$ all $u \in \mathbb{R}^X$ with $u \geqslant 0$

- called the **positive cone** of $\mathbb{R}^X$

An operator $L \in \mathcal{L}(\mathbb{R}^X)$ is called **positive** if

$$u \geqslant 0 \implies Lu \geqslant 0$$

Lemma. $L \in \mathcal{L}(\mathbb{R}^X)$ is positive if and only if its matrix representation is nonnegative

Ex. Prove $L \in \mathcal{L}(\mathbb{R}^X)$ is positive if and only if L is order-preserving on $(\mathbb{R}^X, \leqslant)$

$P \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ is called a **Markov operator** on $\mathbb{R}^{\mathsf{X}}$ if P is positive and $P\mathbb{1} = \mathbb{1}$

We let

$$\mathcal{M}(\mathbb{R}^{\mathsf{X}}) := \text{ the set of all Markov operators on } \mathbb{R}^{\mathsf{X}}$$

Ex. Prove: If $P \in \mathcal{M}(\mathbb{R}^{\mathsf{X}})$ and $v \in \mathbb{R}^{\mathsf{X}}$ with $v \gg 0$, then $Pv \gg 0$.

Ex. Given $P \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$, prove that $P \in \mathcal{M}(\mathbb{R}^{\mathsf{X}})$ if and only if φP defined by

$$(\varphi P)(x') = \sum_{x \in \mathsf{X}} P(x, x')\varphi(x) \qquad (x' \in \mathsf{X})$$

is in $\mathcal{D}(\mathsf{X})$ whenever $\varphi \in \mathcal{D}(\mathsf{X})$