ECON-GA 1025 Macroeconomic Theory I

Lecture 2

John Stachurski

Fall Semester 2018

This Lecture

1. Review of deterministic scalar dynamics
2. Dynamic programming — examples and overview
3. First steps towards analysis / fixed point theory

Warm Up Discussion: Simple Dynamics

Example. Solow–Swan growth

1. Agents save some of their current income
2. Savings used to increase capital stock
3. Capital combined with labor to produce output
4. Output is income (wages, rent on capital)
5. Return to step 1

What happens to output / capital / etc. over time?

In the model, output in each period is

$$Y_t = F(K_t, L_t) \qquad (t = 0, 1, 2, \ldots)$$

Here

- $K_t =$ capital
- $L_t =$ labor
- $Y_t =$ output
- F is the aggregate production function

F assumed to be **homogeneous of degree one** (HD1), meaning

$$F(\lambda K, \lambda L) = \lambda F(K, L) \quad \text{for all } \lambda \geqslant 0$$

Examples.

Cobb-Douglas:

$$F(K, L) = AK^{\alpha} L^{1-\alpha}$$

CES:

$$F(K, L) = \gamma \{ \alpha K^{\rho} + (1 - \alpha) L^{\rho} \}^{1/\rho}$$

Closed economy:

$$\text{current domestic investment} = \text{aggregate domestic savings}$$

The savings rate is a positive constant s, so

$$\text{investment} = \text{savings} = sY_t = sF(K_t, L_t)$$

Depreciation means that 1 unit of capital today becomes $1 - \delta$ units next period

Thus, capital stock evolves according to

$$K_{t+1} = sF(K_t, L_t) + (1 - \delta)K_t$$

We simplify $K_{t+1} = sF(K_t, L_t) + (1-\delta)K_t$ as follows

Assume that $L_t =$ some constant L

Set $k_t := K_t / L$ and use HD1 to get

$$k_{t+1} = s\frac{F(K_t, L)}{L} + (1-\delta)k_t$$

$$= sF(k_t, 1) + (1-\delta)k_t$$

Setting $f(k) := F(k, 1)$, the final expression is

$$k_{t+1} = sf(k_t) + (1-\delta)k_t$$

In summary, we can write

$$k_{t+1} = g(k_t) \quad \text{where} \quad g(k) := sf(k) + (1 - \delta)k$$

This kind of equation is called a (scalar) **difference equation**

Question: What are the implied properties of $\{k_t\}$?

More generally, given

- difference equation $x_{t+1} = g(x_t)$
- initial condition x_0,

what are the properties of $\{x_t\}$?

45 Degree Diagrams

Useful for one dimensional dynamic systems

Equally helpful for both linear and nonlinear systems

Let's look at some examples, starting with the difference equation

$$x_{t+1} = g(x_t) \quad \text{when} \quad g(x) = 2 + 0.5x$$

We want to be able to take any x_0 and map out the sequence

$$x_0, \quad x_1 = g(x_0), \quad x_2 = g(x_1), \quad \ldots$$

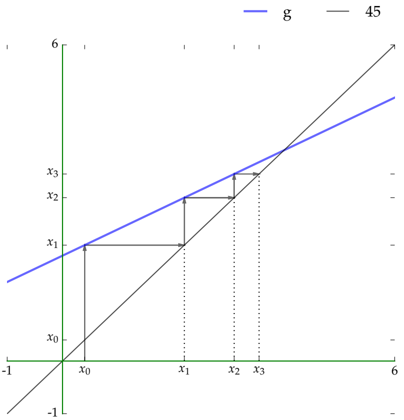


Figure: $g(x) = 2 + 0.5x$ with $x_0 = 0.4$

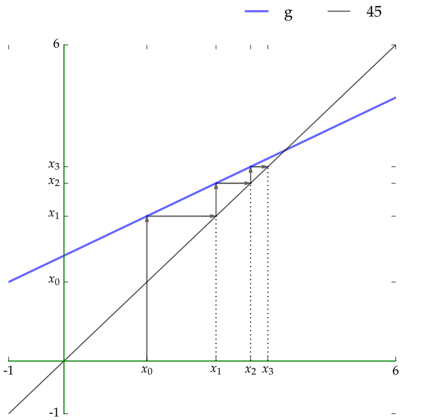


Figure: $g(x) = 2 + 0.5x$ with $x_0 = 1.5$

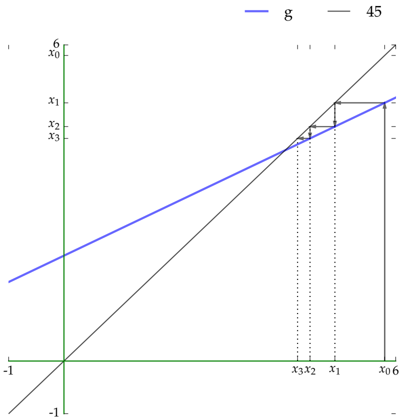


Figure: $g(x) = 2 + 0.5x$ with $x_0 = 5.8$

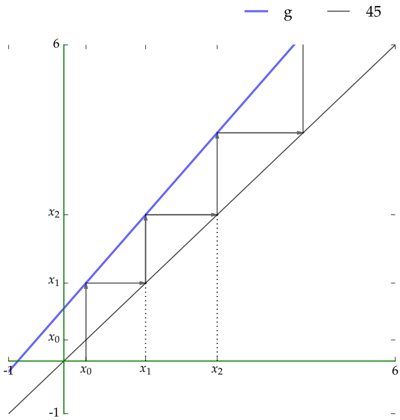


Figure: $g(x) = 1 + 1.2x$ with $x_0 = 0.4$

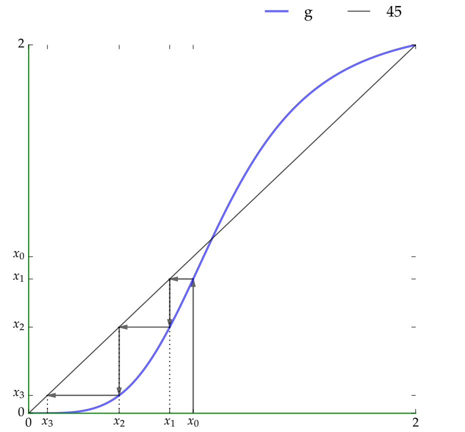


Figure: $g(x) = 2.125/(1 + x^{-4})$ with $x_0 = 0.85$

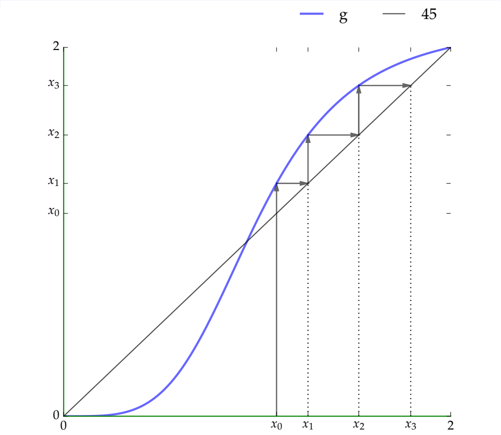


Figure: $g(x) = 2.125/(1 + x^{-4})$ with $x_0 = 1.1$

Let's compare

- 45 degree diagrams
- corresponding time series plots

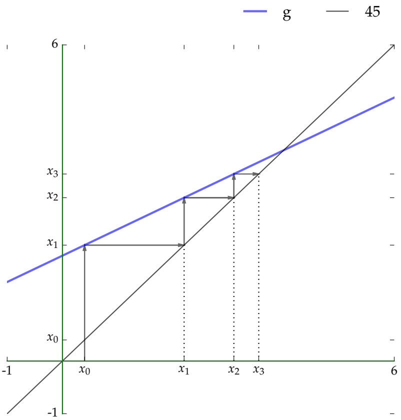


Figure: $g(x) = 2 + 0.5x$ with $x_0 = 0.4$

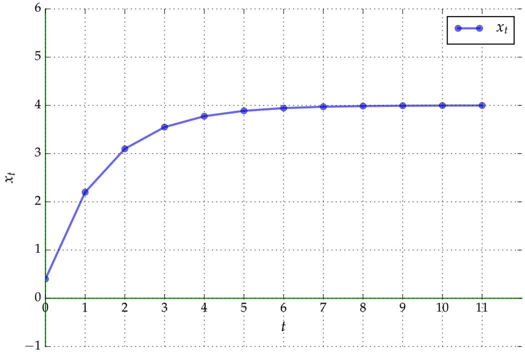


Figure: $g(x) = 2 + 0.5x$ with $x_0 = 0.4$

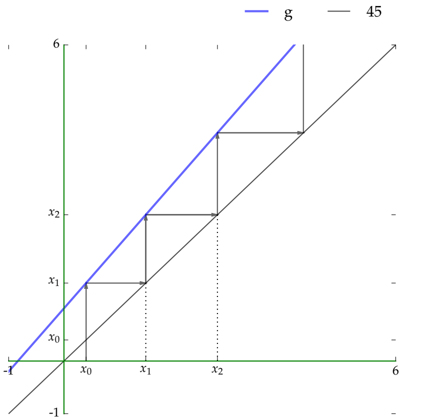


Figure: $g(x) = 1 + 1.2x$ with $x_0 = 0.4$

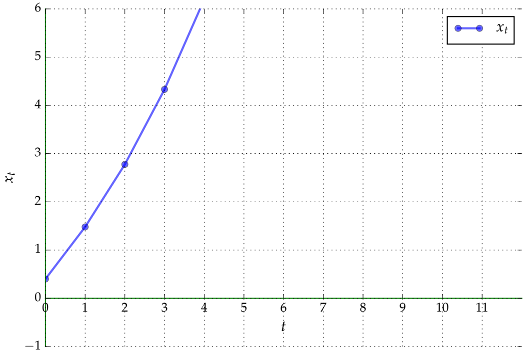


Figure: $g(x) = 1 + 1.2x$ with $x_0 = 0.4$

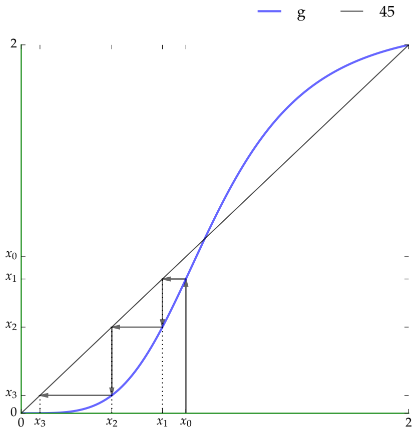


Figure: $g(x) = 2.125/(1 + x^{-4})$ and $g(0) = 0$ with $x_0 = 0.85$

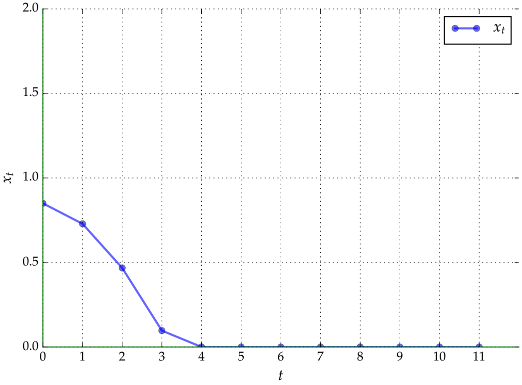


Figure: $g(x) = 2.125/(1 + x^{-4})$ and $g(0) = 0$ with $x_0 = 0.85$

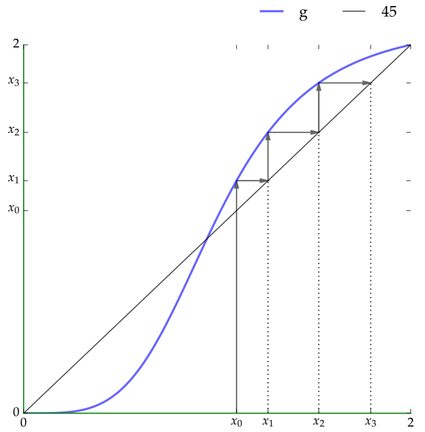


Figure: $g(x) = 2.125/(1 + x^{-4})$ and $g(0) = 0$ with $x_0 = 1.1$

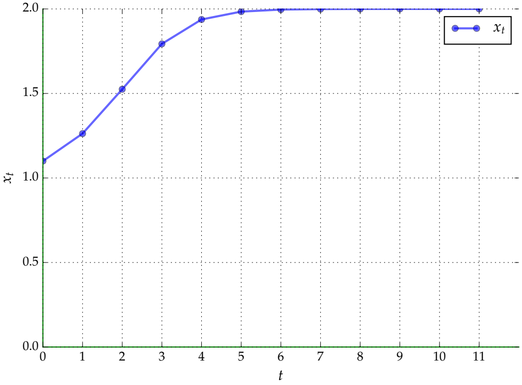


Figure: $g(x) = 2.125/(1 + x^{-4})$ and $g(0) = 0$ with $x_0 = 1.1$

Back to Solow-Swan

Let's return to the model

$$k_{t+1} = g(k_t) \quad \text{where} \quad g(k) := sf(k) + (1 - \delta)k$$

Let's assume that

- $f(k) = Ak^{\alpha}$ where $A = 1$ and $\alpha = 0.6$
- $s = 0.3$ and $\delta = 0.1$

The dynamics can be seen graphically

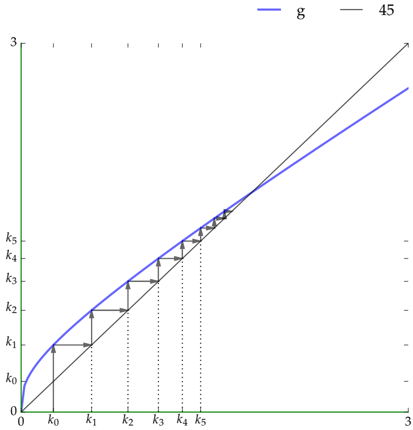


Figure: Solow-Swan dynamics, low initial capital

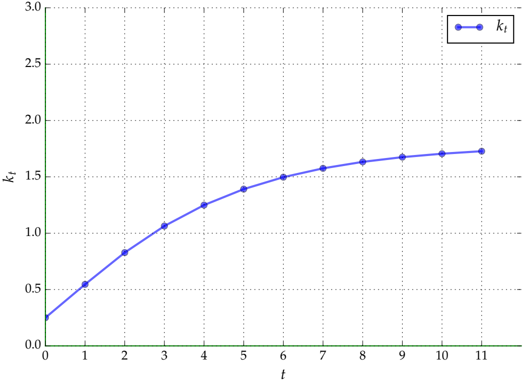


Figure: Solow-Swan dynamics, low initial capital

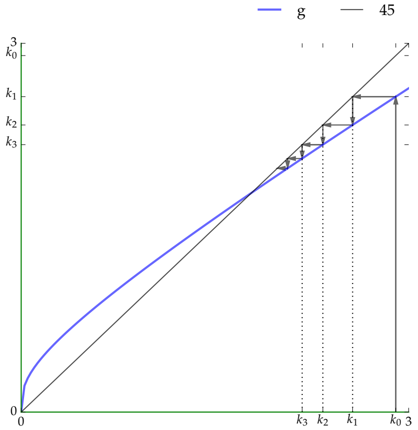


Figure: Solow-Swan dynamics, high initial capital

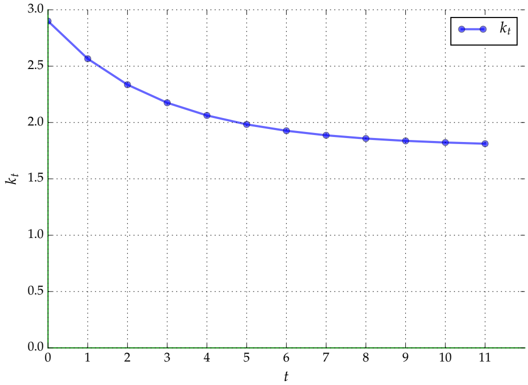


Figure: Solow-Swan dynamics, high initial capital

Graphical analysis of the model suggests that

- k_t increases over time if k_0 is small
- k_t decreases over time if k_0 is large
- k_t converges to the same point regardless of k_0

Adding Complications

Would like to consider random shocks to production, depreciation, etc.

Generates time series in **distribution space**

Tracking them requires some

- functional analysis (distributions are functions)
- numerical methods

Would also like to choose s optimally...

Motivating Examples: Optimization

Some dynamic programming problems

- firm problems
- household problems
- search problems
- etc.

To be solved **in stages** throughout the course

Shortest Paths

A famous topic with applications in

- Google maps!
- operations research
- network design

Aim: traverse a graph, following arcs (arrows) from one specified node to another at minimum cost

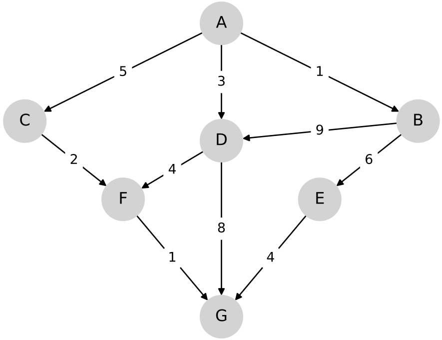


Figure: A simple graph

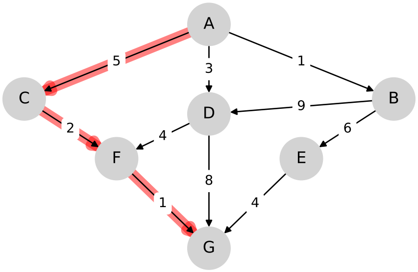


Figure: Solution 1

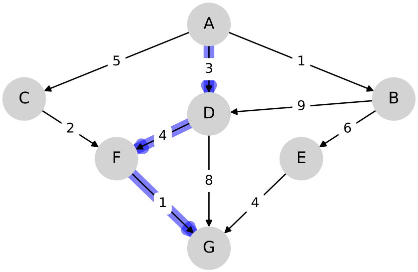


Figure: Solution 2

Large graphs we need a systematic solution

So let $v(x)$ be the **minimum cost-to-go** from node x

The total cost of traveling to the final node from x *if we take the best route*

The function v is usually called the **cost-to-go function** or the **value function**

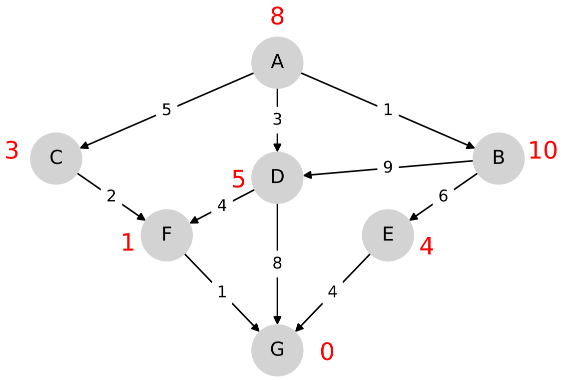


Figure: The cost-to-go function

Suppose that $v(x)$ is known at all nodes x

Then the least cost path can be computed as follows:

Start at node A

From then on, at node x, move to the node y that solves

$$\min_{y \in \Gamma(x)} \{c(x, y) + v(y)\} \tag{1}$$

Here

- $\Gamma(x)$ is the set of nodes that can be reached from x in one step
- $c(x, y)$ is the cost of traveling from x to y

How to find v in more complex cases?

One way is to exploit the recursion

$$v(x) = \min_{y \in \Gamma(x)} \{c(x,y) + v(y)\} \quad \text{for all } x \in \text{graph} \tag{2}$$

Known as the **Bellman equation**

A nonlinear equation in v that we need to figure out how to solve...

Job Search

Let's consider a model of job search due to McCall (1970)

Consider an agent who is currently unemployed

Receives in each period one job offer at wage w_t

On receiving each offer, she has two choices:

1. accept the offer and work permanently at constant wage w_t or
2. reject the offer, receive unemployment compensation c, and reconsider next period

The wage sequence $\{w_t\}$ is assumed to be IID with common density q

Suppose worker enters the workforce at $t = 1$, lives for two periods and maximizes

$$v_1(w_1) := \max\{y_1 + \beta \mathbb{E} y_2\} \quad \text{where } y_j := \text{is income at time } j$$

Income y_j is either wage income or unemployment compensation

Notes

- β lies in $(0, 1)$ and represents discounting of future payoffs
- Smaller β = more impatient
- **Lifetime value** v_1 depends on initial offer w_1

Agent's options:

1. accept w_1 and work at this wage for both periods
2. reject it, receive unemployment compensation c, and then, in the second period, choose the maximum of w_2 and c

Hence

$$v_1(w_1) = \max \{w_1 + \beta w_1, \, c + \beta \, \mathbb{E} \max\{c, w_2\}\} \tag{3}$$

Can be calculated as soon as we know w_1

Now let's suppose that the agent works in period $t = 0$ as well, maximizes

$$v_0(w_0) := \max\{y_0 + \beta \, \mathbb{E} y_1 + \beta^2 \, \mathbb{E} y_2\}$$

The value of accepting the current offer w_0 is $w_0 + \beta w_0 + \beta^2 w_0$

The **continuation value** (i.e., reject, wait) is c plus choosing optimally at $t = 1$ and $t = 2$

Thus,

$$\text{continuation value } = c + \beta \, \mathbb{E} v_1(w_1)$$

We know the function v_1 from the previous slide

Total value from time zero, given w_0, is

$$v_0(w_0) = \max\{\text{accept}, \text{ reject and continue}\}$$

Hence

$$v_0(w_0) = \max\left\{w_0 + \beta\, w_0 + \beta^2\, w_0,\; c + \beta\, \mathbb{E} v_1(w_1)\right\} \tag{4}$$

Note recursive relationship between v_0 and v_1

Also a version of the **Bellman equation**

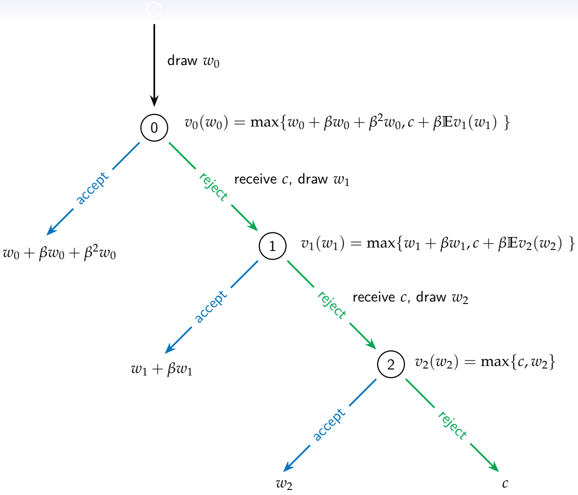


Figure: Decision tree for the job seeker

Now let's suppose that the worker is infinitely lived

Aims to maximize the expected discounted sum

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t y_t \tag{5}$$

The trade-off is

- Waiting for a good offer is costly, since the future is discounted
- Accepting early is costly too, since better offers might arrive

Suppose current wage offer is w

Lifetime value of accepting is

$$w + \beta w + \beta^2 w + \cdots = \frac{w}{1-\beta} \tag{6}$$

Tomorrow we get a random draw w' from q

Let $v^*(w')$ be the **maximum value** that can be extracted from it by making optimal choices at each step

Continuation value is

$$c + \beta \int v^*(w') q(w') \, \mathrm{d}w'$$

Choose the max of these two

But how to find v^*?

The **Bellman equation** states that

$$v^*(w) = \max \left\{ \frac{w}{1-\beta}, \, c + \beta \int v^*(w') q(w') \, \mathrm{d}w' \right\} \tag{7}$$

Intuition: acting optimally today and then continuing to act optimally in the future leads to maximal value today

The Bellman equation is a **restriction** on v^*

We can use it to try to solve for v^*...

Optimal Consumption and Savings

Wealth of a given household evolves according to

$$w_{t+1} = (1 + r_{t+1})(w_t - c_t) + y_{t+1} \tag{8}$$

Here

- w_t is wealth (net asset asset holdings) at t,
- c_t is current consumption,
- y_{t+1} is non-financial (or labor) income received at the end of period t and
- $r_{t+1} > 0$ is the interest rate.

Agent seeks to maximize

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{9}$$

subject to (8) as well as $c_t \geqslant 0$ and $w_t \geqslant 0$ for all t

(Nonnegative wealth excluded at this point)

Here

- $u(c_t)$ is the utility derived from current consumption c_t
- $\beta \in (0, 1)$ is a time discount factor

Assume labor income and the interest rate are functions

$$y_t = y(z_t, \xi_t) \quad \text{and} \quad r_t = r(z_t, \zeta_t) \tag{10}$$

Both ξ_t and ζ_t are **transient shocks**

The sequence $\{z_t\}$ is some **exogenous state process**

It obeys a given transition rule—say

$$z_{t+1} = az_t + b + c\eta_{t+1} \quad \text{with} \quad \{\eta_t\} \overset{\text{IID}}{\sim} N(0,1) \tag{11}$$

Suppose that $v^*(w,z)$ is maximal **lifetime** utility obtainable from wealth w and exogenous state z

We will show: the household should choose c according to

$$\max_{0 \leqslant c \leqslant w} \left\{ u(c) + \beta \, \mathbb{E}_z v^*(w', z') \right\} \tag{12}$$

where

$$w' := (1 + r(z', \xi'))(w - c) + y(z', \zeta')$$

Here $\mathbb{E}_z$ indicates expectation over the random elements $r(z', \xi')$ and $y(z', \zeta')$ conditional on $z_t = z$

But how to find v^*?

Later we show it satisfies

$$v^*(w,z) = \max_{0 \leqslant c \leqslant w} \left\{ u(c) + \beta\, \mathbb{E}_z v^*(w', z') \right\} \tag{13}$$

Intuition: optimally trading of present and future rewards maximizes value

Steps:

1. consider (13) as a functional equation restricting v^*
2. use functional analysis / fixed point theory to solve it

Summary

We will deconstruct high dimensional problems using recursive methods

The recursions lead to **functional equations** like

$$v(w) = \max \left\{ \frac{w}{1-\beta}, \, c + \beta \int v(w') q(w') \, \mathrm{d}w' \right\} \tag{14}$$

or

$$v(w, z) = \max_{0 \leqslant c \leqslant w} \left\{ u(c) + \beta \, \mathbb{E}_z v(w', z') \right\} \tag{15}$$

Unknown v is a function

To solve such equations we use functional analysis / fixed point theory

Next Topics

1. Notational conventions

2. Reminders on real analysis

3. Functional analysis

4. Fixed point theory

Preliminary I: Notation and Conventions

You will see expressions such as $\int g(x)F(\mathrm{d}x)$ where F is a CDF

Interpretation: as

$$\int g(x)F(\mathrm{d}x) = \mathbb{E}g(X) \text{ where } X \overset{\mathscr{D}}{=} F \tag{16}$$

Example. If $g(x) = x$ then $\int g(x)F(\mathrm{d}x)$ is the mean of F

Example. If $g(x) = x^2$ then $\int g(x)F(\mathrm{d}x)$ is the second moment

If X is scalar and $F' = f$, so that f is the density of X, then

$$\int g(x) F(\mathrm{d}x) = \int_{-\infty}^{\infty} g(x) f(x) \, \mathrm{d}x$$

If F corresponds to a PMF p supported on a countable set X, then

$$\int g(x) F(\mathrm{d}x) = \sum_{x \in \mathsf{X}} g(x) p(x)$$

Remarks:

- Lebesgue's theory of integration unifies these concepts
- We skip this topic while borrowing some rules for integrals

Functions on Finite Sets = Vectors

- $\mathbb{R}^d$ is all d-tuples $(x_1, \ldots, x_d)$ of real numbers
- $\mathbb{R}^\mathsf{X}$ is all functions f mapping X to $\mathbb{R}$
  - Each f defined by the value $f(x)$ it assigns to each $x \in \mathsf{X}$

Observe: If $\mathsf{X} = \{x_1, \ldots, x_d\}$ then

$$\mathbb{R}^\mathsf{X} \ni f = (f(x_1), \ldots, f(x_d)) \in \mathbb{R}^d \tag{17}$$

This is a **one-to-one correspondence** between $\mathbb{R}^\mathsf{X}$ and $\mathbb{R}^d$

$$\mathbb{R}^d \ni (y_1, \ldots, y_d) =: (f(x_1), \ldots, f(x_d)) = f \in \mathbb{R}^\mathsf{X} \tag{18}$$

Hence, if X has d elements, then we regard $\mathbb{R}^\mathsf{X}$ and $\mathbb{R}^d$ as the **same set** expressed in different ways

Preliminary II: Real Analysis

Recall that $\{x_n\}$ in $\mathbb{R}$ **converges** to $x \in \mathbb{R}$ if

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N} \text{ s.t. } |x_n - x| < \epsilon \text{ whenever } n \geqslant N$$

Rules for sequences: If $\{x_n\}$ and $\{y_n\}$ are sequences in $\mathbb{R}$ with $x_n \to x$ and $y_n \to y$, then

1. $x_n + y_n \to x + y$ and $x_n y_n \to xy$
2. $x_n \leqslant y_n$ for all n implies $x \leqslant y$
3. $\alpha x_n \to \alpha x$ for any $\alpha \in \mathbb{R}$
4. $x_n \vee y_n \to x \vee y$ and $x_n \wedge y_n \to x \wedge y$

In what follows, a nonempty set X is called **countable** if it is

- finite *or*
- can be placed in one-to-one correspondence with $\mathbb{N}$

Example. $\{1,\ldots,n\}$, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, etc.

Any nonempty set X that fails to be countable is called **uncountable**

Example. $\mathbb{R}$, $\mathbb{R}^d$, $(a,b) \subset \mathbb{R}$, etc.

See any text on real analysis

If $f, g \in \mathbb{R}^{\mathsf{X}}$ then $f + g$, αf, fg to be interpreted pointwise

In particular, for all $x \in \mathsf{X}$,

- $(f + g)(x) := f(x) + g(x)$
- $(\alpha f)(x) := \alpha f(x)$
- $(fg)(x) := f(x)g(x)$
- etc.

Similarly, $f \vee g$, $f \wedge g$ defined by

- $(f \vee g)(x) := f(x) \vee g(x) =$ pointwise max
- $(f \wedge g)(x) := f(x) \wedge g(x) =$ pointwise min

Let $\mathsf{X} \subset \mathbb{R}$

A function $f \in \mathbb{R}^{\mathsf{X}}$ is called **continuous at x** if

$$f(x_n) \to f(x) \quad \text{whenever} \quad x_n \to x$$

The function f is **continuous** if continuity holds at all $x \in \mathsf{X}$

Continuity is preserved under standard algebraic manipulations

Examples.

- f, g continuous $\implies$ $f + g$ continuous
- f, g continuous $\implies$ fg continuous
- etc.

Suggestion for proofs: minimize use of $\forall\, \epsilon > 0,\ \exists \ldots$

Example. To show that f, g continuous implies $f + g$ continuous

Pick any $x \in X$ and any $x_n \to x$

Since f is continuous, $f(x_n) \to f(x)$

Since g is continuous, $g(x_n) \to g(x)$

Since limits of sums are sum of limits,

$$f(x_n) + g(x_n) \to f(x) + g(x) \qquad (n \to \infty)$$

Hence $f + g$ is continuous at x

Since x was arbitrary, $f + g$ is continuous on X

Vector Analysis: Preliminaries

As before, $\mathbb{R}^d$ denotes the set of all d vectors $x = (x_1, \ldots, x_d)$

- In matrix algebra, x defaults to column vector

The **Euclidean norm** $\| \cdot \|$ on $\mathbb{R}^d$ is defined by

$$\|x\| := \left(\sum_{i=1}^{d} x_i^2 \right)^{1/2}$$

Interpretation:

- $\|x\|$ represents the "length" of x
- $\|x - y\|$ represents distance between x and y

Fact. For any $\alpha \in \mathbb{R}$ and any $x, y \in \mathbb{R}^d$, the following statements are true:

1. $\|x\| \geqslant 0$ and $\|x\| = 0$ if and only if $x = 0$
2. $\|\alpha x\| = |\alpha| \|x\|$
3. $\|x + y\| \leqslant \|x\| + \|y\|$ (**triangle inequality**)

The Euclidean norm satisfies the **Cauchy-Schwarz inequality**

$$|x'y| \leqslant \|x\| \|y\|$$

(Here $x'y$ is the **inner product** $\sum_{i=1}^d x_i y_i$)

Order

Let x and y be vectors in $\mathbb{R}^d$

We write $x \leqslant y$ if every element is correspondingly ordered

Examples.

$$\begin{pmatrix} 1 \\ -2 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} 1 \\ -2 \end{pmatrix} \nleqslant \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

Letting e_k be the k-th canonical basis vector,

$$x \leqslant y \quad \iff \quad e_k' x \leqslant e_k' y \text{ in } \mathbb{R} \text{ for all } k$$

Ex. Show that $\leqslant$ is a **partial order** on $\mathbb{R}^d$

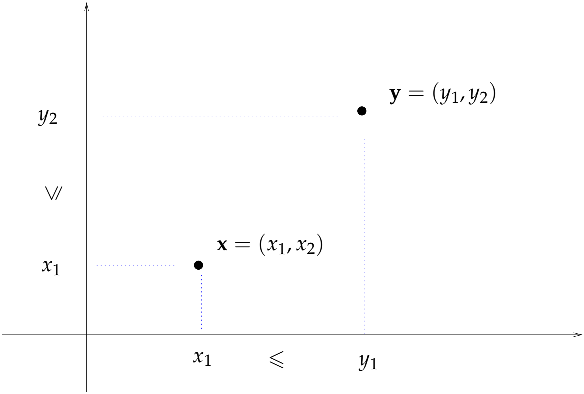


Figure: In $\mathbb{R}^2$, $x \leqslant y$ means y is north-east of x

Sequences and Convergence

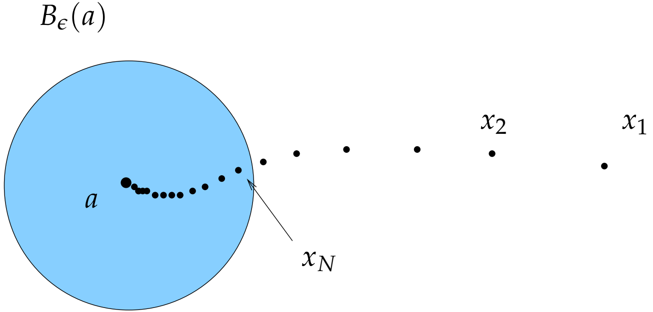
Fix $a \in \mathbb{R}^d$ and $\epsilon > 0$

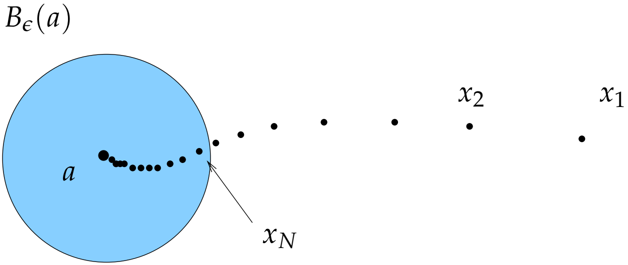
Let $B_\epsilon(a) := \{x \in \mathbb{R}^d : \|x - a\| < \epsilon\}$

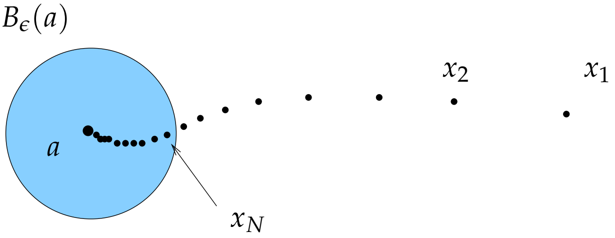
A sequence $\{x_n\}$ said to **converge** to $a \in \mathbb{R}^d$ if

$$\forall \epsilon > 0, \; \exists \, N \in \mathbb{N} \;\; \text{s.t. } n \geqslant \mathbb{N} \implies x_n \in B_\epsilon(a)$$

Equivalent: $\|x_n - a\| \to 0$ in $\mathbb{R}$


$B_\epsilon(a)$
x_2
x_1
a
x_N


$B_\epsilon(a)$
a
x_2
x_1
x_N


$B_\epsilon(a)$
x_2
x_1
a
x_N

Facts Analogous to the scalar case,

1. If $x_n \to x$ and $y_n \to y$ then $x_n + y_n \to x + y$
2. If $x_n \to x$ and $\alpha \in \mathbb{R}$ then $\alpha x_n \to \alpha x$
3. If $x_n \to x$ and $z \in \mathbb{R}^d$ then $z' x_n \to z' x$
4. If $x_n \to x$, $y_n \to y$ and $x_n \leqslant y_n$ for all $n \in \mathbb{N}$, then $x \leqslant y$
5. Each sequence in $\mathbb{R}^d$ has at most one limit

Infinite Sums in $\mathbb{R}^d$

Analogous to the scalar case, an infinite sum in $\mathbb{R}^d$ is the limit of the partial sum:

- If $\{x_n\}$ is a sequence in $\mathbb{R}^d$, then

$$\sum_{n=1}^{\infty} x_n := \lim_{J\to\infty} \sum_{n=1}^{J} x_n \text{ if the limit exists}$$

In other words,

$$y = \sum_{n=1}^{\infty} x_n \quad \iff \quad \lim_{J\to\infty} \left\| \sum_{n=1}^{J} x_n - y \right\| \to 0$$

The Set of Matrices $\mathcal{M}(n \times k)$

Let $\mathcal{M}(n \times k)$ be the set of $n \times k$ real matrices

Questions:

- When is matrix A "close" to matrix B?
- When does A_n converge to A?
- What does $\sum_{n=1}^{\infty} A_n$ mean?

To answer these questions, we introduce a norm on $\mathcal{M}(n \times k)$

The Spectral Norm

Given $A \in \mathcal{M}(n \times k)$, the **spectral norm** of A is

$$\|A\| := \sup \left\{ \frac{\|Ax\|}{\|x\|} \; : \; x \in \mathbb{R}^k, \; x \neq 0 \right\}$$

- LHS is the spectral norm of A
- RHS is ordinary Euclidean vector norms

We often just say the **norm** of A

Fact. In the sup we can restrict attention to x s.t. $\|x\| = 1$

Fact. For the diagonal matrix

$$D = \operatorname{diag}(d_1, \ldots, d_n) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$$

we have

$$\|D\| = \max_i |d_i|$$

Properties of the Spectral Norm

Similar to Euclidean norms on vectors,

Fact. For all $A, B \in \mathcal{M}(n \times k)$,

1. $\|A\| \geqslant 0$ and $\|A\| = 0 \iff A = 0$
2. $\|\alpha A\| = |\alpha| \|A\|$ for any scalar α
3. $\|A + B\| \leqslant \|A\| + \|B\|$

Ex. Show that

$$\|Ax\| \leqslant \|A\| \cdot \|x\| \qquad \forall\, x \in \mathbb{R}^k$$

Fact. If AB is well defined, then $\|AB\| \leqslant \|A\|\|B\|$

Proof: Let $A \in \mathcal{M}(n \times k)$, let $B \in \mathcal{M}(k \times j)$ and let $x \in \mathbb{R}^j$

We have

$$\|ABx\| \leqslant \|A\| \cdot \|Bx\| \leqslant \|A\| \cdot \|B\| \cdot \|x\|$$

$$\therefore \quad \frac{\|ABx\|}{\|x\|} \leqslant \|A\| \cdot \|B\|$$

Called the **submultiplicative property**

Implication: $\|A^j\| \leqslant \|A\|^j$ for any $j \in \mathbb{N}$ and $A \in \mathcal{M}(n \times n)$

Distance, Convergence, etc.

Having a norm on matrices gives us a notion of distance:

$$d(A, B) = \|A - B\|$$

We say that A_j **converges** to A if $\|A_j - A\| \to 0$ in $\mathbb{R}$

Similarly,

$$\sum_{j=1}^{\infty} A_j = B \quad \iff \quad \lim_{J \to \infty} \left\| \sum_{j=1}^{J} A_j - B \right\| = 0$$

A scalar $\lambda \in \mathbb{C}$ is called an **eigenvalue** of $A \in \mathcal{M}(n \times n)$ if there exists a nonzero $e \in \mathbb{C}^n$ such that

$$Ae = \lambda e$$

The vector e is called the **eigenvector**

Ex. A square matrix is called **stochastic** if it is nonnegative and its rows sum to one. Show that 1 is an eigenvalue of every stochastic matrix.

Fact. For any square matrix A

$$\lambda \text{ is an eigenvalue of } A \iff \det(A - \lambda I) = 0$$

Proof: Let A by $n \times n$ and let I be the $n \times n$ identity

We have

$$\begin{aligned}
\det(A - \lambda I) = 0 &\iff A - \lambda I \text{ is singular} \\
&\iff \exists\, x \neq 0 \text{ s.t. } (A - \lambda I)x = 0 \\
&\iff \exists\, x \neq 0 \text{ s.t. } Ax = \lambda x \\
&\iff \lambda \text{ is an eigenvalue of } A
\end{aligned}$$

Example. In the 2×2 case,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Longrightarrow \quad A - \lambda I = \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$$

$$\begin{aligned} \therefore \quad \det(A - \lambda I) &= (a - \lambda)(d - \lambda) - bc \\ &= \lambda^2 - (a + d)\lambda + (ad - bc) \end{aligned}$$

Hence the eigenvalues of A are given by the two roots of

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0$$

Equivalently,

$$\lambda^2 - \text{trace}(A)\lambda + \det(A) = 0$$

Spectral Radius

Let $\sigma(A)$ be the **spectrum** of A (i.e., the set of its eigenvalues)

For $A \in \mathcal{M}(n \times n)$, the **spectral radius** is

$$r(A) := \max_{\lambda \in \sigma(A)} |\lambda|$$

Example. For the diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$ we have

$$\|D\| = \max_i |d_i| = \max_{\lambda \in \sigma(A)} |\lambda| = r(A)$$

Fact. $r(A) \leqslant \|A\|$ always holds

Fact. If A is a stochastic matrix then $r(A) = 1$

Fact. If $a \in \mathbb{R}$ and $A \in \mathcal{M}(n \times n)$, then $r(aA) = |a| r(A)$

Fact. For all $A \in \mathcal{M}(n \times n)$, we have

1. $\|A\| = \sqrt{r(A'A)}$
2. $\|A'\| = \|A\|$ and $r(A') = r(A)$

Gelfand's formula states that, for all $A \in \mathcal{M}(n \times n)$,

$$\|A^k\|^{1/k} \to r(A) \quad \text{as} \quad k \to \infty$$

Ex. Use Gelfand's formula to show that

$$r(A) < 1 \implies \|A^k\| \to 0$$

Proof that $\|A^k\| = O(r(A)^k)$ when A is diagonalizable

Fix $k \in \mathbb{N}$ and P, D such that $A = PDP^{-1}$ where

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

We have $A^k = PD^kP^{-1}$ where $D^k = \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)$

Hence $\|A^k\| = \|PD^kP^{-1}\| \leqslant \|P\|\|D^k\|\|P^{-1}\|$

With $C := \|P\|\|P^{-1}\|$,

$$\|A^k\| \leqslant C \max_i |\lambda_i^k| = C \max_i |\lambda_i|^k = Cr(A)^k$$