MAU34804

Lecture 14

2026-02-24

The Banach fixed point theorem

(Banach Fixed Point Theorem) *Suppose d is a complete metric on a non-empty set X, $f : X \to X$ is a function and $c < 1$ is such that*

$$d(f(\mathbf{w}), f(\mathbf{x})) \le c d(\mathbf{w}, \mathbf{x})$$

for all $\mathbf{w}, \mathbf{x} \in X$. Then f has exactly one fixed point.

This is also called the *contraction mapping principle*.

For $X = \mathbf{R}^n$ define $f(\mathbf{x}) = \mathbf{x} + \mathbf{a}$. This satisfies all the hypotheses above, except with $c = 1$ rather than $c < 1$.

The uniqueness part of the conclusion fails if $\mathbf{a} = \mathbf{0}$ and the existence part fails if $\mathbf{a} \neq \mathbf{0}$.

Note that the hypothesis that f is Lipschitz continuous implies it is uniformly continuous and hence continuous with respect to the metric d.

If X is a subset of $\mathbf{R}^n$ and d is one of the metrics d_p considered earlier then these notions are all independent of which p is chosen, but the Lipschitz constant, which appears in the hypotheses of the theorem, *does* depend on the value of p.

In our applications we will have f's which satisfy the hypotheses for $p = 1$ but not for $p = 2$, so we need non-Euclidean metrics, even for subsets of $\mathbf{R}^n$.

Proof

Uniqueness is easy. Suppose both $\mathbf{w}$ and $\mathbf{x}$ are fixed points. Then

$$(1-c)d(\mathbf{w},\mathbf{x}) = d(\mathbf{w},\mathbf{x}) - cd(\mathbf{w},\mathbf{x}) \leq d(\mathbf{w},\mathbf{x}) - d(f(\mathbf{w}), f(\mathbf{x})) = d(\mathbf{w},\mathbf{x}) - d(\mathbf{w},\mathbf{x}) = 0.$$

From this, $1-c>0$ and $d(\mathbf{w},\mathbf{x}) \geq 0$ it follows that $d(\mathbf{w},\mathbf{x}) = 0$ and hence $\mathbf{w}=\mathbf{x}$.
Existence isn't too much harder.
Suppose $\mathbf{x}_0 \in X$ and define $\mathbf{x}_j$ for $j>0$ recursively by $\mathbf{x}_j = f(\mathbf{x}_{j-1})$.
By induction on l we have $d(\mathbf{x}_l, \mathbf{x}_{l+1}) \leq c^l d(\mathbf{x}_0, \mathbf{x}_1)$.
Then by induction on k and the triangle inequality we have

$$d(\mathbf{x}_j, \mathbf{x}_{j+k}) \leq \frac{c^j - c^{j+k}}{1-c} d(\mathbf{x}_0, \mathbf{x}_1).$$

Indeed the case $k=0$ is trivial and if we assume the inequality holds for a given k then

$$\begin{aligned} d(\mathbf{x}_j, \mathbf{x}_{j+k+1}) &\leq d(\mathbf{x}_j, \mathbf{x}_{j+k}) + d(\mathbf{x}_{j+k}, \mathbf{x}_{j+k+1}) \\ &\leq \frac{c^j - c^{j-k}}{1-c} d(\mathbf{x}_0, \mathbf{x}_1) + c^{j+k} d(\mathbf{x}_0, \mathbf{x}_1) \leq \frac{c^j - c^{j+k+1}}{1-c} d(\mathbf{x}_0, \mathbf{x}_1). \end{aligned}$$

Proof, continued

From

$$d(\mathbf{x}_j, \mathbf{x}_{j+k}) \leq \frac{c^j - c^{j+k}}{1-c} d(\mathbf{x}_0, \mathbf{x}_1)$$

it follows that

$$d(\mathbf{x}_j, \mathbf{x}_{j+k}) \leq \frac{c^j}{1-c} d(\mathbf{x}_0, \mathbf{x}_1).$$

For any $\epsilon > 0$ the right hand side is less than ϵ once

$$j > \frac{d(\mathbf{x}_0, \mathbf{x}_1))}{\epsilon \log(1/c)}.$$

So $\mathbf{x}_0, \mathbf{x}_1, \ldots$ is a Cauchy sequence and, since d is complete, a convergent sequence.
Let $\mathbf{x}^*$ be its limit.
From $\mathbf{x}_j = f(\mathbf{x}_{j-1})$ and the continuity of f we get $\mathbf{x}^* = f(\mathbf{x}^*)$, i.e. that $\mathbf{x}^*$ is a fixed point of f.

Markov Processes

Consider a random variable X which can take any of finitely many values $x_1, x_2, \ldots, x_n$ at times $t_0, t_1, \ldots$.
It follows from the definition of conditional probability that

$$p(X(t_{i+1}) = x_k) = \sum_{j=1}^{n} p(X(t_{i+1}) = x_k \mid X(t_i) = x_j) \cdot p(X(t_i) = x_j).$$

Similarly, $p(X(t_{i+2}) = x_l) = \sum_{j=1}^{n} p(X(t_{i+2}) = x_l \mid X(t_i) = x_j) \cdot p(X(t_i) = x_j)$.
The conditional probability above can be calculated by considering all possible intermediate states $X(t_{i+1})$.

$$\begin{aligned} &p(X(t_{i+2}) = x_l \mid X(t_i) = x_j) \\ &\qquad = \sum_{k=1}^{n} p(X(t_{i+2}) = x_l \mid X(t_{i+1}) = x_k, X(t_i) = x_j) \cdot p(X(t_{i+1}) = x_k \mid X(t_i) = x_j). \end{aligned}$$

Markov processes, continued

The conditional probability $p(X(t_{i+2}) = x_l \,|\, X(t_{i+1}) = x_k, X(t_i) = x_j)$ may or may not depend on j. If it doesn't then it must be equal to $p(X(t_{i+2}) = x_l \,|\, X(t_{i+1}) = x_k)$. Then the last equation on the previous slide simplifies to

$$p(X(t_{i+2}) = x_l \,|\, X(t_i) = x_j) \\ = \sum_{k=1}^{n} p(X(t_{i+2}) = x_l \,|\, X(t_{i+1}) = x_k) \cdot p(X(t_{i+1}) = x_k \,|\, X(t_i) = x_j),$$

from which we get

$$p(X(t_{i+2}) = x_l) \\ = \sum_{j=1}^{n} \sum_{k=1}^{n} p(X(t_{i+2}) = x_l \,|\, X(t_{i+1}) = x_k) \cdot p(X(t_{i+1}) = x_k \,|\, X(t_i) = x_j) \cdot p(X(t_i) = x_j)$$

We can also get this last equation without the assumption that the conditional probability $p(X(t_{i+2}) = x_l \,|\, X(t_{i+1}) = x_k, X(t_i) = x_j)$ is independent of j.

A subtle point

We've proved the relations

$$p(X(t_{i+2}) = x_l) = \sum_{j=1}^{n} p(X(t_{i+2}) = x_l \mid X(t_i) = x_j) \cdot p(X(t_i) = x_j),$$

$$\begin{aligned} &p(X(t_{i+2}) = x_l) \\ &\quad = \sum_{j=1}^{n} \sum_{k=1}^{n} p(X(t_{i+2}) = x_l \mid X(t_{i+1}) = x_k) \cdot p(X(t_{i+1}) = x_k \mid X(t_i) = x_j) \cdot p(X(t_i) = x_j) \end{aligned}$$

without the assumption that the conditional probability $p(X(t_{i+2}) = x_l \mid X(t_{i+1}) = x_k, X(t_i) = x_j)$ is independent of j.
Can we use this to remove that assumption from our earlier proof of

$$\begin{aligned} &p(X(t_{i+2}) = x_l \mid X(t_i) = x_j) \\ &\quad = \sum_{k=1}^{n} p(X(t_{i+2}) = x_l \mid X(t_{i+1}) = x_k) \cdot p(X(t_{i+1}) = x_k \mid X(t_i) = x_j)? \end{aligned}$$

A subtle point, continued

At first glance it looks like we should be able to remove the assumption that $p(X(t_{i+2}) = x_l \,|\, X(t_{i+1}) = x_k, X(t_i) = x_j)$ is independent of j from our proof of the formula relating conditional probabilities, since the $p(X(t_i) = x_j)$ could be (nearly) anything.

Unfortunately this is not true. Consider the case where $n = 3$, $x_i = i$ and $X(t_i)$ is chosen by rolling a die to select one of the six permutations of $\{1, 2, 3\}$ and applying that permutation to the terms in the sequence $1, 2, 3, 1, 2, 3, \ldots$.

It's easy to check that $p(X(t_{i+2}) = l \,|\, X(t_i) = j)$ is 0 if $j = l$ and 1/2 otherwise. Similarly, $p(X(t_{i+2}) = l \,|\, X(t_{i+1}) = k)$ is 0 if $k = l$ and 1/2 otherwise and $p(X(t_{i+1}) = k \,|\, X(t_i) = j)$ is 0 if $j = k$ and 1/2 otherwise.

These numbers however do not satisfy

$$\begin{aligned} &p(X(t_{i+2}) = x_l \,|\, X(t_i) = x_j) \\ &\qquad = \sum_{k=1}^{n} p(X(t_{i+2}) = x_l \,|\, X(t_{i+1}) = x_k) \cdot p(X(t_{i+1}) = x_k \,|\, X(t_i) = x_j). \end{aligned}$$

Markov processes, definition

We say that a stochastic process of the kind just considered is a *Markov process* if it satisfies the following two conditions.
- The conditional probability $p(X(t_{i+1}) = k \,|\, X(t_i) = j)$ is independent of i.
- The conditional probability $p(X(t_{i+1}) = k \,|\, X(t_i) = j, \ldots)$ where $\ldots$ indicates statements about values of X at times before t_i is independent of those values.

The first condition says that while the values of X are time-dependent the underlying process which generates them is time-independent.
The second condition, which is a generalisation of our earlier assumption, says that this process has no memory, in the sense that knowledge of past values is of no help in predicting future values from the current value.

Is economic expansion/contraction a Markov process?

I took US seasonally adjusted quarterly growth data for 1947-2025 and looked at $p(x|yz)$, the observed conditional probability that the next quarter is of type x if the current quarter is of type y and the previous one was of type z, where x, y and z are E for expansionary and C for contractionary, to see if the conditional probabilities are independent of z, as they should be for a Markov process.
I would have used Irish data if I could find it.

$$p(C|CC) = \frac{2}{2+5} = 0.29 \quad p(C|CE) = \frac{5}{5+7} = 0.42$$

$$p(C|EC) = \frac{2}{2+10} = 0.17 \quad p(C|EE) = \frac{10}{10+272} = 0.04$$

$$p(E|CC) = \frac{5}{2+5} = 0.71 \quad p(E|CE) = \frac{7}{5+7} = 0.58$$

$$p(E|EC) = \frac{10}{2+10} = 0.83 \quad p(E|EE) = \frac{272}{10+272} = 0.96$$