Introduction to Chaotic Dynamics and Fractals

Abbas Edalat
ae@ic.ac.uk

Imperial College London

Bertinoro, June 2013

Topics covered

- Discrete dynamical systems
- Periodic doublig route to chaos
- Iterated Function Systems and fractals
- Attractor neural networks

Continuous maps of metric spaces

- We work with metric spaces, usually a subset of $\mathbb{R}^n$ with the Euclidean norm or the space of code sequences such as $\Sigma^{\mathbb{N}}$ with an appropriate metric.
- A map of metric spaces $F : X \to Y$ is **continuous at** $x \in X$ if it preserves the limits of convergent sequences, i.e., for all sequences $(x_n)_{n\geq 0}$ in X:

$$x_n \to x \Rightarrow F(x_n) \to F(x).$$

- F is **continuous** if it is continuous at all $x \in X$.
- **Examples**:
  - All polynomials, $\sin x$, $\cos x$, e^x are continuous maps.
  - $x \mapsto 1/x : \mathbb{R} \to \mathbb{R}$ is not continuous at $x = 0$ however we define $1/0$. Similarly for $\tan x$ at $x = (n + \frac{1}{2})\pi$ for any integer n.
  - The step function $s : \mathbb{R} \to \mathbb{R} : x \mapsto 0$ if $x \leq 0$ and 1 otherwise, is not continuous at 0.
  - Intuitively, a map $\mathbb{R} \to \mathbb{R}$ is continuous iff its graph can be drawn with a pen without leaving the paper.

Continuity and Computability

- Continuity of F is necessary for the computability of F.
- Here is a simple argument for $F : \mathbb{R} \to \mathbb{R}$ to illustrate this.
- An irrational number like π has an infinite decimal expansion and is computable only as the limit of an effective sequence of rationals $(x_n)_{n\geq 0}$ with say $x_0 = 3, x_1 = 3.1, x_2 = 3.14 \cdots$.
- Hence to compute $F(\pi)$ our only hope is to compute $F(x_n)$ for each rational x_n and then take the limit. This requires $F(x_n) \to F(\pi)$ as $n \to \infty$.

Discrete dynamical systems

- A **discrete dynamical system** $F : X \to X$ is the action of a continuous map F on a metric space (X, d), usually a subset of $\mathbb{R}^n$.
- Here are some key continuous maps giving rise to interesting dynamical systems in $\mathbb{R}^n$:
- Linear maps $\mathbb{R}^n \to \mathbb{R}^n$, eg $x \mapsto ax : \mathbb{R} \to \mathbb{R}$ for any $a \in \mathbb{R}$.
- The quadratic family $F_c : \mathbb{R} \to \mathbb{R} : x \mapsto cx(1 - x)$ for any $c \in [1, 4]$.

Differential equations

- Differential equations are continuous dynamical systems which can be studied using discrete dynamical systems.
- Let $\dot{y} = V(y) \in \mathbb{R}^n$ be a system of differential equations in $\mathbb{R}^n$ with initial condition $y(0) = x_0$ at $t = 0$.
- Suppose a solution of the system at time t is $y(t) = S(x_0, t)$.
- Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be given by $F(x) = S(x, 1)$.
- Then, F is the **time-one map** of the evolution of the differential equation with $y(0) = F^0(x_0)$, $y(1) = F(x_0)$, $y(2) = F(F(x_0))$, $y(3) = F(F(F(x_0)))$ and so on.
- By choosing the unit interval of time, we can then study the solution to the differential equation by studying the discrete system F.

Iteration

- Given a function $F : X \to X$ and an initial value x_0, what ultimately happens to the sequence of iterates
$$x_0, F(x_0), F(F(x_0)), F(F(F(x_0))), \ldots.$$
- We shall use the notation
$$F^{(2)}(x) = F(F(x)), F^{(3)}(x) = F(F(F(x))), \ldots$$
For simplicity, when there is no ambiguity, we drop the brackets in the exponent and write
$$F^n(x) := F^{(n)}(x).$$
- Thus our goal is to describe the *asymptotic behaviour* of the iteration of the function F, i.e. the behaviour of $F^n(x_0)$ as $n \to \infty$ for various initial points x_0.

Orbits

Definition

Given $x_0 \in X$, we define the **orbit of x_0 under F** to be the sequence of points

$$x_0 = F^0(x_0), x_1 = F(x_0), x_2 = F^2(x_0), \ldots, x_n = F^n(x_0), \ldots.$$

The point x_0 is called the **seed** of the orbit.

Example

If $F(x) = \sin(x)$, the orbit of $x_0 = 123$ is

$$x_0 = 123, x_1 = -0.4599..., x_2 = -0.4439..., x_3 = -0.4294...,$$

Finite Orbits

Definition

- A **fixed point** is a point x_0 that satisfies $F(x_0) = x_0$.
- A fixed point x_0 gives rise to a **constant orbit**: $x_0, x_0, x_0, \ldots$.
- The point x_0 is **periodic** if $F^n(x_0) = x_0$ for some $n > 0$. The least such n is called the **period** of the orbit. Such an orbit is a repeating sequence of numbers.
- A point x_0 is called **eventually fixed** or **eventually periodic** if x_0 itself is not fixed or periodic, but some point on the orbit of x_0 is fixed or periodic.

Graphical Analysis

Given the graph of a function F we plot the orbit of a point x_0.

- First, superimpose the diagonal line $y = x$ on the graph. (The points of intersection are the fixed points of F.)
- Begin at (x_0, x_0) on the diagonal. Draw a vertical line to the graph of F, meeting it at $(x_0, F(x_0))$.
- From this point draw a horizontal line to the diagonal finishing at $(F(x_0), F(x_0))$. This gives us $F(x_0)$, the next point on the orbit of x_0.
- Draw another vertical line to graph of F, intersecting it at $F^2(x_0))$.
- From this point draw a horizontal line to the diagonal meeting it at $(F^2(x_0), F^2(x_0))$.
- This gives us $F^2(x_0)$, the next point on the orbit of x_0.
- Continue this procedure, known as **graphical analysis**. The resulting "staircase" visualises the orbit of x_0.

Graphical analysis of linear maps

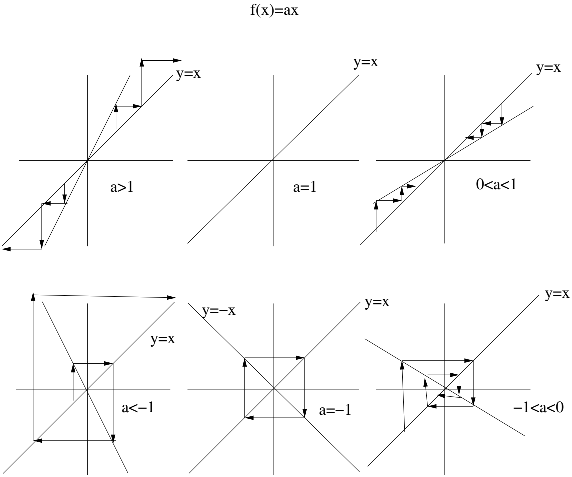
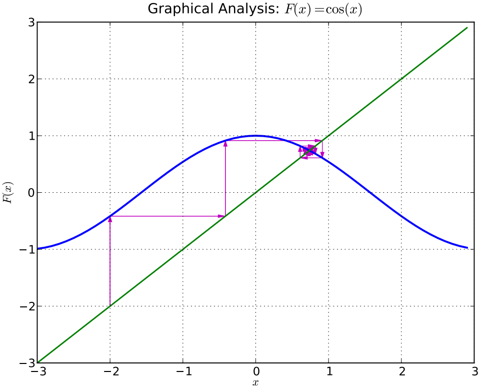


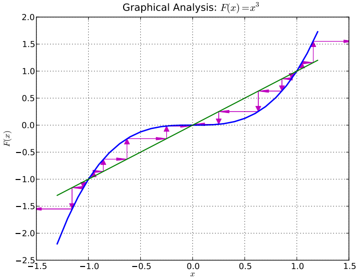
Figure : Graphical analysis of $x \mapsto ax$ for various ranges of $a \in \mathbb{R}$.

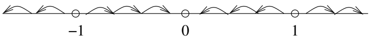
A Non-linear Example: $C(x) = \cos x$



Phase portrait

- Sometimes we can use graphical analysis to describe the behaviour of *all* orbits of a dynamical system.
- In this case we say that we have performed a complete **orbit analysis** which gives rise to the **phase portrait** of the system.
- **Example:** The complete orbit analysis of $x \mapsto x^3$ and its phase portrait are shown below.





Phase portraits of linear maps

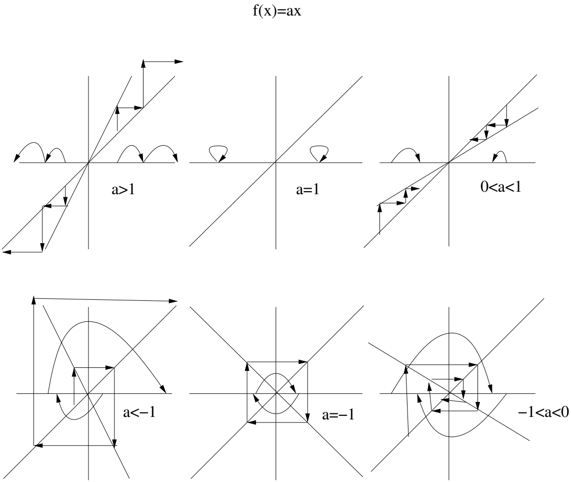


Figure : Graphical analysis of $x \mapsto ax$ for various ranges of $a \in \mathbb{R}$.

Open and closed subsets

- Given a metric space (X, d), the **open ball** with centre $x \in X$ and radius $r > 0$ is the subset $O(x, r) = \{y \in X : d(x, y) < r\}$.
- Eg, in $\mathbb{R}$, if $a < b$, then the interval $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ is an open ball; it is called an open interval.
- **An open set** O is any union of open balls: $O = \bigcup_{i \in I} O(x_i, r_i)$, where I is any indexing set.
- A **closed set** is the complement of an open set.
- Eg, in $\mathbb{R}$, if $a \leq b$, then the interval $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ is closed.
- $[a, b) = \{x : a \leq x < b\}$ is neither open nor closed.

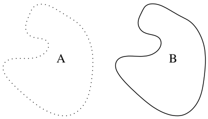


Figure : An open and a closed set

Properties of Open and closed subsets

- The following properties follow directly from the definition of open and closed sets in any metric space (X, d).
- X and the empty set $\emptyset$ are both open and closed.
- An arbitrary union of open sets is open while an arbitrary intersection of closed sets is closed.
- Furthermore, any finite intersection of open sets is open while any finite union of closed sets is closed.
- Note that even countable intersection of open sets may not be open, eg.

$$\bigcap_{n\geq 1}(0, 1 + \frac{1}{n}) = (0, 1].$$

Open subsets and continuity

- Suppose $F : X \to Y$ is a map of metric spaces.
- Given $B \subset Y$, the **pre-image** of B under F is the set

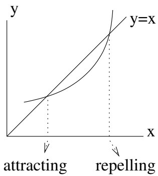
$$F^{-1}(B) = \{x \in X : F(x) \in B\}.$$

- It can be shown that given a map of metric spaces $F : X \to Y$ and $x \in X$, then the following are equivalent:
  - F is continuous at $x \in X$ (i.e., it preserves the limit of convergent sequences).
  - $\forall \epsilon > 0. \exists \delta > 0$ such that $F[O(x, \delta)] \subset O(f(x), \epsilon)$, (equivalently $O(x, \delta) \subset F^{-1}(O(f(x), \epsilon))$).
- $F : X \to Y$ is continuous (i.e., it is continuous at every point of X) iff the pre-image of any open set in Y is open in X.

Attracting and repelling periodic points

A set B is **invariant** under F if $F(x) \in B$ if $x \in B$.

Suppose x_0 is a periodic point for F with period n. Then x_0 is an **attracting periodic point** if for $G = F^n$ the orbits of points in some invariant open neighbourhood of x_0 converge to x_0. The point x_0 is a **repelling periodic point** if for $G = F^n$ the orbits of all points in some open neighbourhood of x_0 (with the exception of the trivial orbit of x_0) eventually leave the neighbourhood.



It can be shown that if F is differentiable and its derivative F' is continuous at a fixed point x_0 of F, then x_0 is attracting (repelling) if $|F'(x_0)| < 1$ ($|F'(x_0)| > 1$). If $|F'(x_0)| \neq 1$, then x_0 is called a **hyperbolic** fixed point.

Attractors

- We have already seen periodic attractors which are finite sets.
- Attractors can generally be very complex sets.
- Recall that an **open set** in a metric space (X, d) is any union of open balls of the form $O(x, r) = \{y \in X : d(x, y) < r\}$ where $x \in X$ and $r > 0$.
- Given $F : X \to X$, we say a non-empty closed subset $A \subset X$ is an **attractor** if it satisfies:
  - Closure under iteration: $x \in A \Rightarrow F(x) \in A$
  - There is a **basin** of attraction for A, i.e., there exists an invariant open set $B \subset X$ with $A \subset B$ such that for any $x \in B$ and any open neighbourhood O of A, there exists $N \in \mathbb{N}$ such that $F^n(x) \in O$ for all $n \geq N$.
  - No proper subset of A satisfies the above two properties.
- Attractors are observable by computing iterates of maps.

Periodic attractors

Attracting periodic points give rise to attractors.

- Here is a proof:
- Let $x \in X$ be an attracting period point of $F : X \to X$ with period $n \geq 1$.
- Claim: $A = \{x, F(x), F^2(x), \cdots, F^{n-1}(x)\}$ is an attractor.
- It is clearly non-empty and closed as any finite set is closed.
- By definition there is an open neighbourhood $O \subset X$ of x that consists of points whose orbits under the map F^n remain in O, i.e., $F^n[O] \subset O$, and converge to x.
- Let $B = O \cup F^{-1}(O) \cup F^{-2}(O) \cup \cdots \cup F^{-(n-1)}(O)$.
- Then, B is an open subset which satisfies $F[B] \subset B$ (since $F[F^{-i}(O)] \subset F^{-(i-1)}(O)$ for $1 \leq i \leq n-1$) and is a basin of attraction for A.
- Since $F^n(x) = x$ with n minimal, A is not decomposable.

First observed “strange” attractor: Lorenz attractor

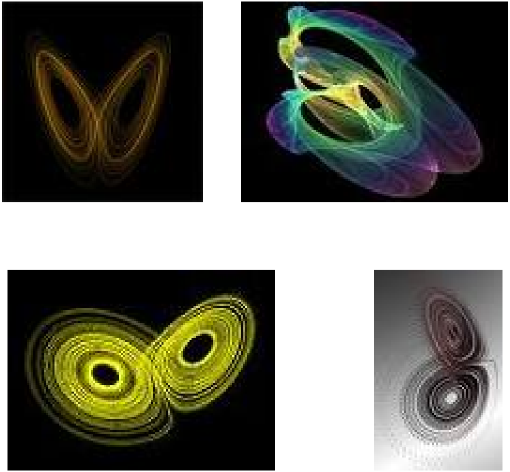


Figure : Different viewpoints of Lorenz attractor in $\mathbb{R}^3$

It has been recently proved that the Lorenz attractor is **chaotic.**

Chaos

Definition

A dynamical system $f : X \to X$ on an infinite metric space X is **sensitive to initial conditions** if there exists $\delta > 0$ such that for any $x \in X$ and any open ball O with centre x, there exists $y \in O$ and $n \in \mathbb{N}$ such that $d(f^n(x), f^n(y)) > \delta$ (i.e., in the neighbourhood of any point x there is a point whose iterates eventually separate from those of x by more than δ).

$f : X \to X$ is **chaotic** if it satisfies the following:

(i) f is sensitive to initial conditions.

(ii) f is **topologically transitive**, i.e., for any pair of open balls $U, V \subset X$, there exists $n > 0$ such that $f^n(U) \cap V \neq \emptyset$, i.e., f is not decomposable.

(iii) The periodic orbits of f form a dense subset of X, i.e., any open ball contains a periodic point.

It can be shown that (i) follows from (ii) and (iii).

Space of infinite strings

- Let $\Sigma = \{0, 1\}$.
- Consider the set $(\Sigma^{\mathbb{N}}, d)$ or $(\Sigma^{\mathbb{Z}}, d)$ of infinite strings $x = x_0 x_1 x_2 \cdots$ or $x = \cdots x_{-2} x_{-1} x_0 x_1 x_2 \cdots$, where $x_k \in \Sigma$, with metric:

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1/2^{|m|} & |m| \in \mathbb{N} \text{ least with } x_m \neq y_m. \end{cases}$$

Here are some simple properties to check:

- $d(x, y)$ is clearly non-negative, symmetric and vanishes iff $x = y$.
- d also satisfies the triangular inequality, hence is a metric.
- The open ball with centre $x \in \Sigma^{\mathbb{N}}$ and radius $1/2^n$ for $n \geq 0$ is:
$$O(x, 1/2^n) = \{y : \forall i \in \mathbb{Z}.\, |i| \leq n \Rightarrow \; x_i = y_i\}.$$

Tail map is chaotic

- The tail map $\sigma : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ is defined by

  $$(\sigma(x))_n = x_{n+1}$$

  for $n \in \mathbb{N}$.
- The tail map is also called the **shift** map as it shifts all components one place to the left.
- The tail map is continuous and satisfies the following:
  - Sensitive to initial conditions (with $\delta = 1/2$).
  - Topologically transitive ($\forall$ open balls U. $\exists n.\ \sigma^n(U) = \Sigma^{\mathbb{N}}$).
  - Its periodic orbits are dense in $\Sigma^{\mathbb{N}}$
  - It has a dense orbit.
- In particular, σ is chaotic.

Turing machines

- A Turing machine is a dynamical system (Y, T) as follows.
- Let $Y = \{0,1\}^{\mathbb{Z}} \times S$ where S is a finite set of *states*, which contains an element 0 called the halting state.
- The set $\{(\ldots, 0, \ldots)\} \times S$ is called the empty tape.
- The infinite sequences with only finite number of 1 in $\{0,1\}^{\mathbb{Z}}$ are called data.
- The Turing machine is defined by three maps:

$f : \{0,1\} \times S \to \{0,1\}$	defines the new letter
$g : \{0,1\} \times S \to S$	defines the new state
$h : \{0,1\} \times S \to \{-1,0,1\}$	decides to move left, right or stay

- Now we define $T : \{0,1\}^{\mathbb{Z}} \times S \to \{0,1\}^{\mathbb{Z}} \times S$:

$$T(x,s) = \left(\sigma^{h(x_0,s)}(\ldots, x_{-2}, x_{-1}, f(x_0,s), x_1, x_2, \ldots), g(x_0,s)\right)$$

Angle doubling is chaotic

Let S^1 be the unit circle centred at the origin, whose points are angles $0 \leq \theta < 2\pi$ measured in radians. The distance between any two points on the circle is the shortest arc between them.

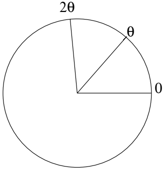


Figure : Angle doubling map

- Angle doubling: $A_2 : S^1 \to S^1$ is defined as $A_2(\theta) = 2\theta$.
- It can be shown that A_2 satisfies the following:
  - Sensitive to initial conditions (with $\delta = \pi/4$).
  - Topologically transitive ($\forall$ open balls U. $\exists n$. $A_2^n(U) = S^1$).
  - Its periodic orbits are dense in S^1.
- **Theorem.** The angle doubling map is chaotic.

Topological Conjugacy

Suppose $f : X \to X$ and $g : Y \to Y$ are continuous maps and $h : X \to Y$ is a surjective continuous map such that $h \circ f = g \circ h$:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] & X \arrow[d, "h"] \\ Y \arrow[r, "g"] & Y \end{tikzcd}$$

- We then say h is a **topological semi-conjugacy** between (X, f) and (Y, g). Note: $h \circ f^n = g^n \circ h$ for any $n \in \mathbb{N}$, for $n = 2$:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] & X \arrow[r, "f"] \arrow[d, "h"'] & X \arrow[d, "h"] \\ Y \arrow[r, "g"] & Y \arrow[r, "g"] & Y \end{tikzcd}$$

- h maps orbits to orbits, periodic points to periodic points.
- If h has a continuous inverse, then it is called a **topological conjugacy** between (X, f) and (Y, g).

Example of topological conjugacy

- Consider $Q_d : \mathbb{R} \to \mathbb{R} : x \mapsto x^2 + d$
- For $d < 1/4$, the map Q_d is conjugate via a linear map of type $L : x \mapsto \alpha x + \beta$ to $F_c : \mathbb{R} \to \mathbb{R} : x \mapsto cx(1 - x)$ for a unique $c > 1$.
- This can be shown by finding α, β, c in terms of d such that the conjugacy relation holds, i.e.,:

$$L^{-1} F_c L = Q_d$$

Symbolic dynamics

- Suppose $g : Y \to Y$ is a dynamical system with a semi-conjugacy:

$$\begin{tikzcd} \Sigma^{\mathbb{N}} \arrow[r, "\sigma"] \arrow[d, "h"'] & \Sigma^{\mathbb{N}} \arrow[d, "h"] \\ Y \arrow[r, "g"] & Y \end{tikzcd}$$

- We say $\sigma : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ provides **symbolic dynamics** for $g : Y \to Y$.
- From such a semi-conjugacy we can deduce the following results about g from the corresponding properties of σ:
  - g is topologically transitive.
  - Periodic orbits of g are dense.
  - g has a dense orbit.

An Example

- For $x \in \mathbb{R}$ and integer $n \geq 1$, define $x \bmod n$ as the remainder of division of x by n.
- Thus, $x \bmod 1$ is simply the fractional part of x.
- Consider the map $M_2 : [0, 1) \to [0, 1) : x \mapsto 2x \bmod 1$.
- Then we have the semi-conjugacy $h : \Sigma^{\mathbb{N}} \to [0, 1)$ given by

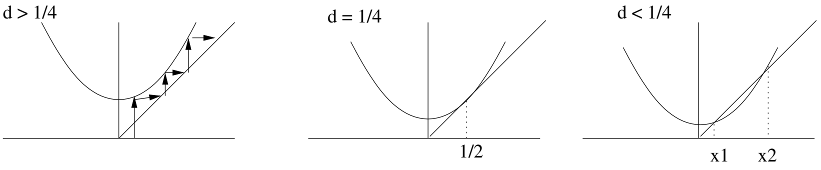
$$h(x) = \Sigma_{i=0}^{\infty} \frac{x_i}{2^i} \bmod 1$$

- It follows immediately that
  - M_2 is topologically transitive.
  - Periodic orbits of M_2 are dense.
  - M_2 has a dense orbit.
- Since M_2 expands any small interval by a factor 2, it is also sensitive to initial conditions.
- Therefore, M_2 is a chaotic map.

Periodic doubling bifurcation to chaos

Bifurcation

- Consider the one-parameter family of quadratic maps $x \mapsto x^2 + d$ where $d \in \mathbb{R}$.
- For $d > 1/4$, no fixed points and all orbits tend to ∞.
- For $d = 1/4$, a fixed point at $x = 1/2$, the double root of $x^2 + 1/4 = x$.
- This fixed point is locally attracting on the left $x < 1/2$ and repelling on the right $x > 1/2$.
- For d just less than $1/4$, two fixed points $x_1 < x_2$, with x_1 attracting and x_2 repelling.
- The family $x \mapsto x^2 + d$ undergoes **bifurcation** at $d = 1/4$.



A model in population dynamics

- Consider the one-parameter family $F_c : [0,1] \to [0,1]$ with $c > 1$, given by $F_c(x) = cx(1-x)$.
- Think of $x \in [0,1]$ as the fraction of predator in a predator-prey population.
- F_c gives this fraction after one unit of time, where the predator population increases linearly $x \mapsto cx$ but it is also impeded because of limitation of resources, i.e., prey, by the quadratic factor $-cx^2$.
- The dynamical system $F_c : \mathbb{R} \to \mathbb{R}$ is studied for $c > 1$.
- It turns out that this simple system exhibits highly sophisticated including chaotic behaviour.

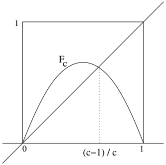


Figure : The quadratic family

Nature of fixed points

- There are two fixed points at 0 and $p_c = (c-1)/c < 1$.
- $F'(x) = c(1-2x)$, so $F'_c(0) = c > 1$ and $F'_c(p_c) = 2 - c$.
- 0 is always a repeller and p_c is an attractor for $1 < c < 3$ and a repeller for $c > 3$.
- For $1 < c < 3$ the orbit of any point in $(0, 1)$ tends to p_c.
- So what happens when c passes through 3 turning p_c from an attracting fixed point to a repelling one?
- How do we find the phase portrait of F_c for c just above 3?

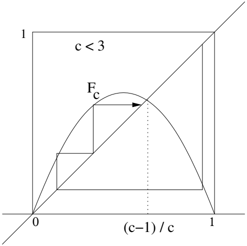


Figure : Dynamics for $c < 3$

Period doubling

- We look at the behaviour of F_c^2 when c passes through 3.
- For $c < 3$, the map F_c^2 has (like F_c) two fixed points, one repelling at 0 and one attracting at p_c .
- As c passes through 3, the fixed point p_c becomes repelling for F_c^2. (We have: $F_3'(p_3) = -1$ and $(F_3^2)'(p_3) = 1$)
- However, F_c^2 acquires two new attracting fixed points a_c, b_c with $a_c < p_c < b_c$ for c just above 3.
- This gives an attracting period 2 orbit $\{a_c, b_c\}$ for F_c.

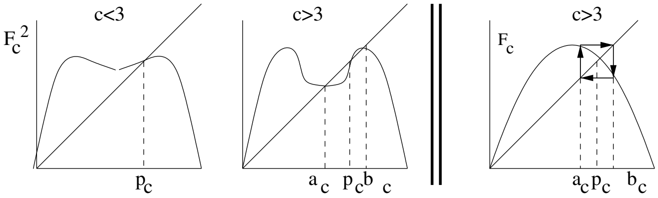


Figure : An attracting period 2 orbit is born

Period doubling bifurcation diagram

- Therefore, F_c goes under a periodic doubling bifurcation as c goes through 3.
- If we plot the attractors of F_c as c passes through $c = 3$ we obtain the following diagram.
- The solid lines indicate attracting fixed or periodic points while the dashed line indicate the repelling fixed point.

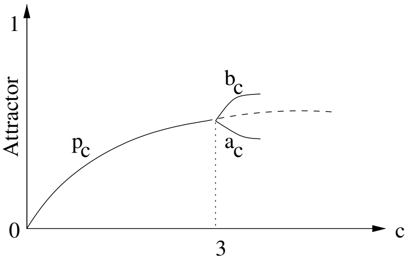


Figure : Periodic doubling bifurcation diagram

Analysis of periodic doubling

- F_c^2 has four fixed points satisfying: $x = F_c^2(x) = c(cx(1-x))(1-cx(1-x))$.
- Two of them are 0 and p_c, the fixed points of F_c.
- The other two namely a_c and b_c are given by:
- $a_c = \frac{c+1-\sqrt{c^2-2c-3}}{2c}$ and $b_c = \frac{c+1+\sqrt{c^2-2c-3}}{2c}$.
- The periodic point a_c (or b_c) will be attracting as long as $|(F_c^2)'(a_c)| = |F_c'(b_c)F_c'(a_c)| < 1$.
- This gives $3 < c < 1+\sqrt{6}$.
- It follows that for c in the above range all orbits, except the orbits that land on 0 and p_c, are in the basin of attraction of the periodic orbit $\{a_c, b_c\}$.
- At $c_1 = 1+\sqrt{6}$, we have $(F_c^2)'(a_c) = (F_c^2)'(b_c) = -1$.

A repeat of period doubling

- The periodic doubling repeats itself as c passes through $c_1 = 1 + \sqrt{6} > c_0 = 3$.
- For $c > c_1$, $(F_c^2)'(a_c) = (F_c^2)'(b_c) > 1$ and the periodic orbit $\{a_c, b_c\}$ becomes repelling.
- The attracting periodic points of period 2 namely a_c and b_c now become repelling.
- An attracting periodic orbit of period four is created with the four points close to a_c and b_c as in the diagram.

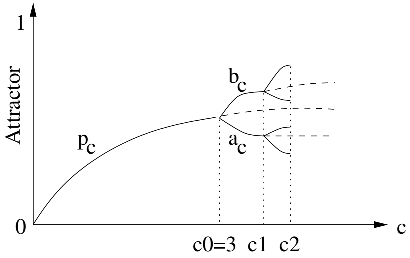


Figure : A second periodic doubling bifurcation

Renormalisation

- Note that F_c at c_0 behaves like F_c^2 at c_1.
- Fixed points cease to be attracting and become repelling.
- This similarity can be understood using **renormalisation**.
- Let $q_c = 1/c$ be the point that is mapped to p_c ($c > 2$).
- Put $R(F_c) = L_c \circ F_c^2 \circ L_c^{-1}$, where L_c is the linear map $x \mapsto (x - p_c)/(q_c - p_c)$, which maps $[q_c, p_c]$ into $[0, 1]$.
- If $F_{c'}(1/2) = R(F_c)(1/2)$ then $F_{c'}$ and $R(F_c)$ are close.

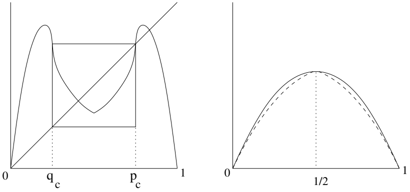


Figure : Left: F_c^2 graph and its renormalising part in the square, Right: $F_{2.24}$ (solid curve) close to $R(F_{3.3})$ (dashed curve)

A cascade of periodic doubling

- The periodic doubling bifurcation repeats itself ad infinitum for an increasing sequence: $c_0 < c_1 < c_2 < c_3 \cdots$.
- As c passes through c_n, for $n \geq 0$, the attracting periodic orbit of period 2^n becomes repelling and an attracting periodic orbit of period 2^{n+1} is created nearby.
- The phase portrait of F_c for $c_n < c \leq c_{n+1}$ is simple:
- The basin of attraction of the periodic orbit of period 2^{n+1} consists of all orbits in $(0, 1)$ with the exception of orbits landing on repelling periodic points of period 2^n.
- We have $\lim_{n\to\infty} c_n = c_\infty \approx 3.569\cdots$.
- At $c = c_\infty$ is the system is at **the edge of chaos.** Chaotic behaviour can be detected for many values in $c_\infty < c \leq 4$.
- We say that c_∞ is a **critical** parameter. The system undergoes **phase transition** from regular behaviour to chaotic behaviour as c passes through c_∞.

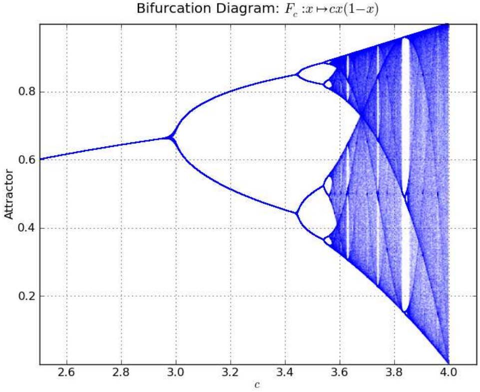


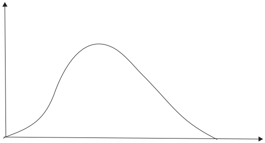
Figure : Bifurcation diagram of F_c: At each value of c, the attractor of F_c is plotted on the vertical line through c. This is done e.g. by computing a typical orbit of F_c, discarding the first 100 iterates and plotting the next 100 iterates. As c increases from 2.5, the family goes through a cascade of periodic doubling bifurcations with periodic attractors of period 2^n for all $n \in \mathbb{N}$, until it reaches the edge of chaos at the critical parameter c_∞. Beyond this value, the family has chaotic behaviour for some values of c interrupted by regular behaviour with attractors of different periods. An attractor of period 3 is clearly visible after $c = 3.8$.

Feigenbaum constant and Unimodal maps

- The following remarkable property holds:

$$\lim_{n\to\infty} \frac{c_n - c_{n-1}}{c_{n+1} - c_n} = \delta \approx 4.66920.$$

- δ is the *Feigenbaum universal constant*.
- In fact, the periodic doubling route to chaos occurs in all the so called *unimodal* families of maps (figure below).
- For example in the family $S_c : [0,1] \to [0,1]$ with $S_c(x) = c \sin \pi x$, with $0 < c < 1$, we discover the same period doubling bifurcation to chaos as we did for the quadratic family.
- Moreover, for all these maps we have $(c_n - c_{n-1})/(c_{n+1} - c_n) \to \delta \approx 4.66920$.



Chaotic dynamics in quadratic family

- We show that F_4 is chaotic.
- The two-to-one map $h_1 : S^1 \to [-1, 1]$ with $h_1(\theta) = \cos\theta$ gives a semi-conjugacy between angle doubling $A_2 : S^1 \to S^1$ and the map $Q : [-1, 1] \to [-1, 1]$ with $Q(x) = 2x^2 - 1$.
- The map $h_2 : [-1, 1] \to [0, 1]$ with $h_2(x) = (1 - x)/2$ gives a conjugacy between Q and F_4.
- Thus, $h_2 \circ h_1 : S^1 \to [0, 1]$ gives a semi-conjugacy between A_2 and F_4.

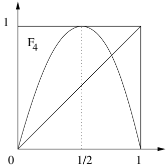


Figure : F_4 map

F_4 is chaotic

- We have the following diagram. Put $g = h_2 \circ h_1$

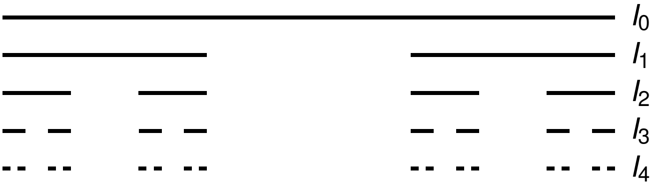
$$\begin{tikzcd}
S^1 \arrow[r, "A_2"] \arrow[d, "h_1"'] & S^1 \arrow[d, "h_1"] \\
{[-1,1]} \arrow[r, "Q"] \arrow[d, "h_2"'] & {[-1,1]} \arrow[d, "h_2"] \\
{[0,1]} \arrow[r, "F_4"] & {[0,1]}
\end{tikzcd}$$

- For any open set $O \subset [0, 1]$, there is some $n \in \mathbb{N}$ with $A_2^n(g^{-1}(O)) = S^1$, which implies $F_4^n(O) = [0, 1]$.
- Sensitive dependence on initial conditions and topological transitivity of F_4 both follow from the above fact.
- Furthermore, $g^{-1}(O) \subset S^1$ is an open set and thus has a periodic point which is mapped by g to a periodic point in O.
- Therefore, F_4 is chaotic.

Iterated Function Systems and Fractals

Cantor middle-thirds set

- Start with $I_0 = [0, 1]$.
- Remove the interior of its middle third to get $I_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$.
- Do the same with each interval in I_1 to get I_2 and so on.
- We have
$$I_{n+1} \;=\; f_1[I_n] \cup f_2[I_n]$$
where the affine transformations $f_1, f_2 : I_0 \to I_0$ are given by: $f_1(x) \;=\; \frac{x}{3}$ and $f_2(x) \;=\; \frac{2}{3} \;+\; \frac{x}{3}$.
- The Cantor set C is the intersection of all I_n's.



Properties of the Cantor set

- C satisfies the fixed point equation: $C = f_1[C] \cup f_2[C]$.
- All the end points of the intervals in I_n are in C.
- C consists of all numbers in the unit interval whose base 3-expansion does not contain the digit 1:

$$\frac{a_0}{3} + \frac{a_1}{3^2} + \frac{a_2}{3^3} + \dots \qquad \text{where } a_i = 0 \text{ or } 2.$$

- C has **fine structure** and **self-similarity** or **scaling invariance**: It is made up of two copies of itself scaled by $\frac{1}{3}$.
- C is closed and uncountable but has zero length.
- Complicated local structure:
  - C is **totally disconnected** (between any two points in C there is a point not in C) and therefore contains no intervals;
  - C has no **isolated point** (in any neighbourhood of a point in C there are infinitely many points of C).

Fractals

- Cantor middle-thirds set is an example of fractals, geometric objects with fine structure, some self-similarity and a fractional fractal dimension.
- A set which can be made up with m copies of itself scaled by $\frac{1}{n}$ has **similarity dimension** $\frac{\log m}{\log n}$, which coincides with its **fractal dimension**.
- The unit interval has dimension 1 as it is precisely the union of n copies of itself each scaled by a factor n.
- The Cantor middle-thirds set has fractal dimension $log2/log3 < 1$.
- More generally, a Cantor set is a closed and totally disconnected set (i.e., contains no proper intervals) without any isolated points (i.e., any neighbourhood of any of its points contains other points of the set).
- A Cantor set may be "fat", i.e., have non-zero "total length".

Chaotic dynamics on fractals

- Let $T : C \to C$ be given by $T(a) = 3a \bmod 1$, where for any real number r its fractional part is denoted by $r \bmod 1$.
- Consider the map $h : \{0,2\}^{\mathbb{N}} \to C$ into the Cantor set with

$$h(x) = \frac{x_0}{3} + \frac{x_1}{3^2} + \frac{x_2}{3^3} \cdots, \text{ where } x = x_0 x_1 x_2 \cdots$$

- We have the topological conjugacy

$$\begin{tikzcd} \{0,2\}^{\mathbb{N}} \arrow[r, "\sigma"] \arrow[d, "h"'] & \{0,2\}^{\mathbb{N}} \arrow[d, "h"] \\ C \arrow[r, "T"] & C \end{tikzcd}$$

- Using this conjugacy, we can show that $T : C \to C$ is chaotic.

Sierpinsky triangle

- The Sierpinsky triangle is another example of a self-similar fractal, generated by three contracting affine maps: $f_1, f_2, f_3 : \mathbb{R}^2 \to \mathbb{R}^2$.
- Each map f_i is a composition of scaling by $1/2$ and a translation.

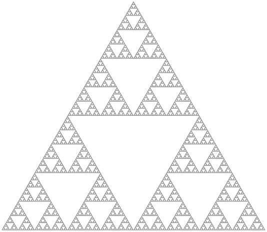
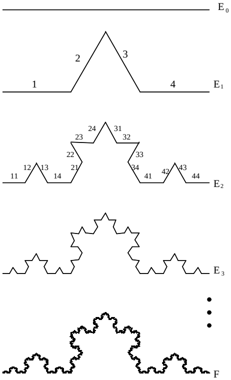


Figure : Sierpinsky triangle

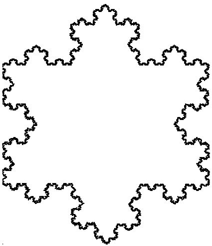
Koch curve

- Start with the unit interval.
- Remove the middle third of the interval and replace it by the other two sides of the equilateral triangle based on the removed segment.
- The resulting set E_1 has four segments; label them 1,2,3,4 from left to right.
- Apply the algorithm to each of these to get E_2.
- Repeat to get E_n for all $n \geq 1$.
- The limiting curve F is the Koch curve, Figure (a) next page.
- Joining three such curves gives us the snowflake as in Figure (b).
- We can map $\{1, 2, 3, 4\}^*$ to the sides of E_n for $n \geq 0$.
- There is a continuous map from $\{1, 2, 3, 4\}^{\mathbb{N}}$ to the Koch curve.

Koch curve



(a)



(b)

Properties of the Koch curve

- F is the limit of a sequence of simple polygons which are recursively defined.
- Fine structure, self-similarity: It is made of four parts, each similar to F but scaled by $\frac{1}{3}$.
- The fractal dimension of the Koch curve is $\log 4 / \log 3$ which is between 1 and 2.
- Complicated local structure: F is nowhere smooth (no tangents anywhere). Think about the sequence of segments $2, 22, 222, \cdots$. It spirals around infinitely many times!
- F has infinite length (E_n has length $(\frac{4}{3})^n$ and this tends to infinity with n) but occupies zero area. The snowflake can be painted but you cannot make a string go around it!
- Although F is defined recursively in a simple way, its geometry is not easily described classically (cf. definition of a circle). But $\{1, 2, 3, 4\}^{\mathbb{N}}$ gives a good model for F.

Affine maps

- An *affine* map of type $\mathbb{R}^m \to \mathbb{R}^m$ is a linear map followed by a translation.
- For example, in $\mathbb{R}^2$, an affine map $f : \mathbb{R}^2 \to \mathbb{R}^2$ has, in matrix notation, the following action:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} k \\ l \end{pmatrix}.$$

- We know how the middle thirds Cantor and the Sierpinsky's triangle are generated by contracting affine maps.
- For the generation of the Koch curve there are four contracting affine transformations of the plane f_i, $1 \leq i \leq 4$, each a combination of a simple contraction, a rotation and a translation, such that

$$E_{n+1} \;=\; f_1[E_n] \cup f_2[E_n] \cup f_3[E_n] \cup f_4[E_n].$$

Iterated Function Systems

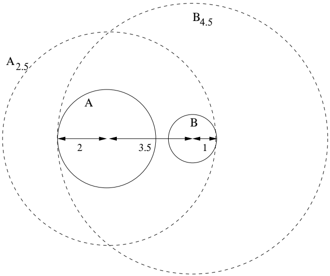
- The examples we have studied are instances of a general system which is a rich source of fractals and we now study.
- An *Iterated Function System (IFS)* in $\mathbb{R}^m$ consists of a finite number of contracting maps $f_i : \mathbb{R}^m \to \mathbb{R}^m$, $i = 1, 2 \ldots N$, i.e, for each i, there exists $0 \le s_i < 1$ such that $\forall.\, x, y \in \mathbb{R}^m$

$$\|f_i(x) - f_i(y)\| \le s_i \|x - y\| \text{ where } \|x\| = \sqrt{\sum_{j=1}^{m} x_j^2}.$$

- If f_i is an affine map then s_i can be taken to be the l_2 norm of its linear part.
- $\mathcal{P}(\mathbb{R}^m)$: set of non-empty, bounded, closed subsets of $\mathbb{R}^m$.
- A continuous map, in particular a contracting map, takes any bounded, closed subset to another such set.
- Let $f : \mathcal{P}(\mathbb{R}^m) \to \mathcal{P}(\mathbb{R}^m) : A \mapsto \bigcup_{1 \le i \le N} f_i[A]$, a well-defined map.
- We will show that f has a unique fixed point, A^* with $f(A^*) = A^*$, which is the **attractor** of the IFS.

Hausdorff distance

- The **Hausdorff distance**, $d_H(A, B)$, between two elements A and B in $\mathcal{P}(\mathbb{R}^m)$ is the infimum of numbers r such that every point of A is within distance r of some point of B and every point of B is within distance r of some point of A.
- Formally, $d_H(A, B) = \inf\{r \mid B \subseteq A_r \text{ and } A \subseteq B_r\}$, where for $r \geq 0$, the *r-neighbourhood* A_r of A is:
$$A_r = \{x \in \mathbb{R}^m \mid \|x - a\| \leq r \text{ for some } a \in A\}.$$
- Eg, if A and B are disks with radius 2 and 1 whose centres are 3.5 units apart, then $d_H(A, B) = 4.5$.



IFS attractor

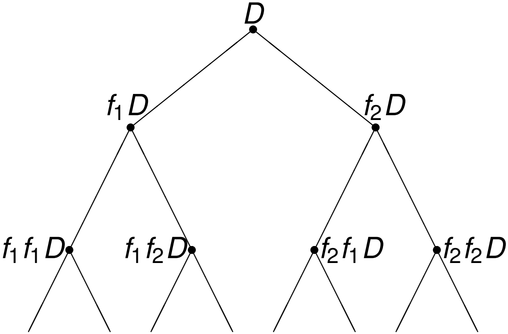
- $(\mathcal{P}(\mathbb{R}^m), d_H)$ is a complete metric space (i.e., every Cauchy sequence has a limit) and the map f is a contracting map wrt d_H with a contractivity $s = \max\{s_i : 1 \leq i \leq N\}$.
- By the contracting mapping theorem, f has a unique fixed point $A^* \in \mathcal{P}(\mathbb{R}^m)$ which is the only attractor of the IFS.
- The attractor is obtained as follows. Let D_{r_i} be the disk with radius $r_i = \|f_i(\mathbf{0})\|/(1 - s_i)$ centred at the origin.
- D_{r_i} is mapped by f_i into itself. (Check this!)
- Put $r = \max_i r_i$. Then $D = D_r$ is mapped by f into itself:

  $$D \supseteq f[D] \supseteq f^2[D] \supseteq \cdots$$

  and $A^* = \bigcap_{n \geq 0} f^n[D]$.
- We can also define the notion of **IFS with probabilities** by assigning a probability weight p_i to each f_i. Then there will be an ergodic measure with **support** A^*.

IFS tree

- The nth iterate $f^n[D] = \bigcup_{i_1, i_2, \cdots i_n = 1}^{N} f_{i_1}[f_{i_2}[\cdots[f_{i_n} D]\cdots]$ generates the N^n nodes of the nth level of the *IFS tree*.
- The N^n nodes on the nth level may have overlaps.
- For $N = 2$, three levels of this tree are shown below.



- Each branch is a sequence of contracting subsets of $\mathbb{R}^m$:

$$D \supseteq f_{i_1} D \supseteq f_{i_1} f_{i_2} D \supseteq f_{i_1} f_{i_2} f_{i_3} D \supseteq \cdots$$

  whose intersection contains a single point.
- A^* is the set of these single points for all the branches.

IFS algorithm

- We use the IFS tree to obtain an algorithm to generate a discrete approximation to the attractor A^* up to a given $\epsilon > 0$ accuracy. In other words, we will obtain a finite set A such that $d_H(A, A^*) \leq \epsilon$.
- The diameter of each node at level n is at most $2rs^n$.
- We need $2rs^n \leq \epsilon$ so that the diameter is at most ϵ.
- Let $n = \left\lceil \frac{\log(\epsilon/2r)}{\log s} \right\rceil$. Consider the truncated tree at level n. The diameters of the N^n leaves of the tree are at most ϵ.
- Pick the distinguished point

$$f_{i_1} f_{i_2} \cdots f_{i_n}(0) \in f_{i_1} f_{i_2} \cdots f_{i_n} D$$

for each leaf. Let A be the set of these N^n points.
- Each point in A^* is in one of the N^n leaves each of which has diameter at most ϵ and contains one of the distinguished points and hence one point of A. It follows that $d_H(A, A^*) \leq \epsilon$ as required.

Complexity of the IFS algorithm

- The complexity of the algorithm is $O(N^n)$.
- This is polynomial in N for a fixed resolution (and thus a fixed n) but exponential in n.
- Improve the efficiency of the algorithm by taking a smaller set of leaves as follows:
- For each branch

  $$D \supset f_{i_1} D \supset f_{i_1} f_{i_2} D \supset f_{i_1} f_{i_2} f_{i_3} D \ldots$$

  of the tree find an integer k such that the diameter of $f_{i_1} \cdots f_{i_k} D$ is at most ϵ by taking the first integer k such that $2rs_{i_1} s_{i_2} \cdots s_{i_k} \leq \epsilon$.
- Then take this node as a leaf. Do as before with this new set of leaves.

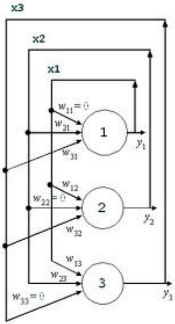
Attractor Neural Networks

The Hopfield network I

- In 1982, John Hopfield introduced an artificial neural network to store and retrieve memory like the human brain.
- Here, a neuron either is **on** (firing) or is **off** (not firing), a vast simplification of the real situation.
- The state of a neuron (on: +1 or off: -1) will be renewed depending on the input it receives from other neurons.
- A Hopfield network is initially trained to store a number of patterns or memories.
- It is then able to recognise any of the learned patterns by exposure to only partial or even some corrupted information about that pattern, i.e., it eventually settles down and returns the closest pattern or the best guess.
- Thus, like the human brain, the Hopfield model has stability in pattern recognition.

The Hopfield network II

- A Hopfield network is single-layered and **recurrent** network: the neurons are fully connected, i.e., every neuron is connected to every other neuron.
- Given two neurons i and j there is a connectivity weight w_{ij} between them which is symmetric $w_{ij} = w_{ji}$ with zero self-connectivity $w_{ii} = 0$.
- Below three neurons $i = 1, 2, 3$ with values ± 1 have connectivity w_{ij}; any update has input x_i and output y_i.



Updating rule

- Assume N neurons $i = 1, \cdots, N$ with values $x_i = \pm 1$
- $X = \{-1, 1\}^N$ with the Hamming distance: $H(\vec{x}, \vec{y}) = \#\{i : x_i \neq y_i\}$.
- The update rule is:

$$\text{If } h_i \geq 0 \text{ then } 1 \leftarrow x_i \text{ otherwise } -1 \leftarrow x_i$$

  where $h_i = \sum_{j=1}^{N} w_{ij} x_j + b_i$ and $b_i \in \mathbb{R}$ is a bias.
- We put $b_i = 0$ for simplicity as it makes no difference to training the network with random patterns but the results we present all extend to $b_i \neq 0$.
- We therefore assume $h_i = \sum_{j=1}^{N} w_{ij} x_j$.
- There are now two ways to update the nodes:
- **Asynchronously:** At each point in time, update one node chosen randomly or according to some rule.
- **Synchronously:** Every time, update all nodes together.
- Asynchronous updating is more biologically realistic.

A simple example

- Suppose we only have two neurons: $N = 2$.
- Then there are essentially two non-trivial choices for connectivities (i) $w_{12} = w_{21} = 1$ or (ii) $w_{12} = w_{21} = -1$.
- **Asynchronous updating:** In the case of (i) there are two attracting fixed points namely $[1, 1]$ and $[-1, -1]$. All orbits converge to one of these. For (ii), the attracting fixed points are $[-1, 1]$ and $[1, -1]$ and all orbits converge to one of these. Therefore, in both cases, the network is **sign blind**: for any attracting fixed point, swapping all the signs gives another attracting fixed point.
- **Synchronous updating:** In both cases of (i) and (ii), although there are fixed points, none attract nearby points, i.e., they are not attracting fixed points. There are also orbits which oscillate forever.

Energy function

- Hopfield networks have an energy function such that every time the network is updated asynchronously the energy level decreases (or is unchanged).
- For a given state (x_i) of the network and for any set of connection weights w_{ij} with $w_{ij} = w_{ji}$ and $w_{ii} = 0$, let

$$E = -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij}x_i x_j$$

- We update x_m to x'_m and denote the new energy by E'.
- **Exercise:** Show that $E' - E = \sum_{i \neq m} w_{mi}x_i(x_m - x'_m)$.
- Using the above equality, if $x_m = x'_m$ then we have $E' = E$.
- If $x_m = -1$ and $x'_m = 1$, then $x_m - x'_m = -2$ and $\sum_i w_{mi}x_i \geq 0$. Thus, $E' - E \leq 0$.
- Similarly if $x_m = 1$ and $x'_m = -1$, then $x_m - x'_m = 2$ and $\sum_i w_{mi}x_i < 0$. Thus, $E' - E < 0$.

Neurons pull in or push away each other

- Consider the connection weight $w_{ij} = w_{ji}$ between two neurons i and j.
- If $w_{ij} > 0$, the updating rule implies:
  - when $x_j = 1$ then the contribution of j in the weighted sum, i.e. $w_{ij}x_j$, is positive. Thus x_i is pulled by j towards its value $x_j = 1$;
  - when $x_j = -1$ then $w_{ij}x_j$, is negative, and x_i is again pulled by j towards its value $x_j = -1$.
- Thus, if $w_{ij} > 0$, then i is pulled by j towards its value. By symmetry j is also pulled by i towards its value.
- If $w_{ij} < 0$ however, then i is pushed away by j from its value and vice versa.
- It follows that for a given set of values $x_i \in \{-1, 1\}$ for $1 \leq i \leq N$, the choice of weights taken as $w_{ij} = x_i x_j$ for $1 \leq i \leq N$ corresponds to the Hebbian rule: "Neurons that fire together, wire together. Neurons that fire out of sync, fail to link."

Training the network: one pattern ($b_i = 0$)

- Suppose the vector $\vec{x} = (x_1, \ldots, x_i, \ldots, x_N) \in \{-1, 1\}^N$ is a pattern we like to store in the memory of a Hopfield network.
- To construct a Hopfield network that remembers $\vec{x}$, we need to choose the connection weights w_{ij} appropriately.
- If we choose $w_{ij} = \eta x_i x_j$ for $1 \leq i, j \leq N$ ($i \neq j$), where $\eta > 0$ is the learning rate, then the values x_i will not change under updating as we show below.
- We have

$$\sum_{j=1}^{N} w_{ij} x_j = \eta \sum_{j \neq i} x_i x_j x_j = \eta \sum_{j \neq i} x_i = \eta (N-1) x_i$$

- This implies that the value of x_i, whether 1 or -1 will not change, so that $\vec{x}$ is a fixed point.
- Note that $-\vec{x}$ also becomes a fixed point when we train the network with $\vec{x}$ confirming that Hopfield networks are sign blind.

Training the network: Many patterns

- More generally, if we have p patterns $\vec{x}^k$, $k = 1, \ldots, p$, we choose $w_{ij} = \frac{1}{N}\sum_{k=1}^{p} x_i^k x_j^k$.
- This is called the generalized Hebbian rule.
- We will have a fixed point $\vec{x}^k$ for each k iff $\mathrm{sgn}(h_i^k) = x_i^k$ for all $1 \leq i \leq N$, where

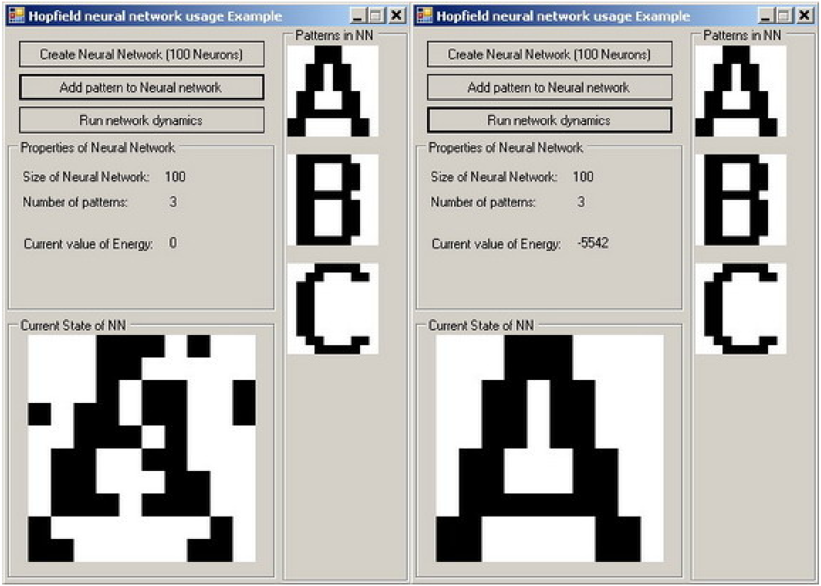
$$h_i^k = \sum_{j=1}^{N} w_{ij} x_j^k = \frac{1}{N} \sum_{j=1}^{N} \sum_{\ell=1}^{p} x_i^\ell x_j^\ell x_j^k$$

- Split the above sum to the case $\ell = k$ and the rest:

$$h_i^k = x_i^k + \frac{1}{N} \sum_{j=1}^{N} \sum_{\ell \neq k} x_i^\ell x_j^\ell x_j^k$$

- If the second term, called the **crosstalk term**, is less than one in absolute value for all i, then h_i^k will not change and pattern k will become a fixed point.
- In this situation every pattern $\vec{x}^k$ becomes a fixed point and we have an associative or content-addressable memory.

Pattern Recognition



Stability of the stored patterns

- How many random patterns can we store in a Hopfield network with N nodes?
- In other words, given N, what is an upper bound for p, the number of stored patterns, such that the crosstalk term remains less than one with high probability?
- Multiply the crosstalk term by $-x_i^k$ to define:

$$C_i^k := -x_i^k \frac{1}{N} \sum_{j=1}^{N} \sum_{\ell \neq k} x_i^\ell x_j^\ell x_j^k$$

- If C_i^k is negative, then the crosstalk term has the same sign as the desired x_i^k and thus this value will not change.
- If, however, C_i^k is positive and greater than 1, then the sign of h_i will change, i.e., x_i^k will change, which means that node i would become unstable.
- We will estimate the probability that $C_i^k > 1$.

Distribution of C_i^k

- For $1 \leq i \leq N$, $1 \leq \ell \leq p$ with both N and p large, consider x_i^ℓ as purely random with equal probabilities 1 and -1.
- Thus, C_i^k is $1/N$ times the sum of (roughly) Np independent and identically distributed (i.i.d.) random variables, say y_m for $1 \leq m \leq Np$, with equal probabilities of 1 and -1.
- Note that $\langle y_m \rangle = 0$ with variance $\langle y_m^2 - \langle y_m \rangle^2 \rangle = 1$ for all m.
- **Central Limit Theorem:** If z_m is a sequence of i.i.d. random variables each with mean μ and variance σ^2 then for large n

$$X_n = \frac{1}{n} \sum_{m=1}^{n} z_m$$

  has approximately a normal distribution with mean $\langle X_n \rangle = \mu$ and variance $\langle X_n^2 - \langle X_n \rangle^2 \rangle = \sigma^2/n$.
- Thus for large N, the random variable $p(\frac{1}{Np}\sum_{m=1}^{Np} y_m)$, i.e., C_i^k, has approximately a normal distribution $\mathcal{N}(0, \sigma^2)$ with mean 0 and variance $\sigma^2 = p^2(1/(Np)) = p/N$.

Storage capacity

- Therefore if we store p patterns in a Hopfield network with a large number of N nodes, then the probability of error, i.e., the probability that $C_i^k > 1$, is:

$$P_{\text{error}} = P(C_i^k > 1) \approx \frac{1}{\sqrt{2\pi}\sigma} \int_1^\infty \exp(-x^2/2\sigma^2)\, dx$$

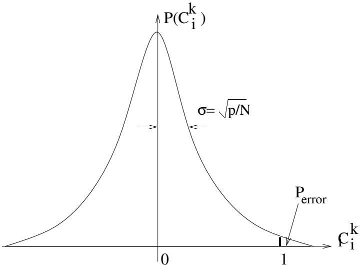
$$= \frac{1}{2}(1 - \text{erf}(1/\sqrt{2\sigma^2})) = \frac{1}{2}(1 - \text{erf}(\sqrt{N/2p}))$$

where the **error** function erf is given by:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-s^2)\, ds.$$

- Therefore, given N and p we can find out what the probability P_{error} of error is.

Storage capacity



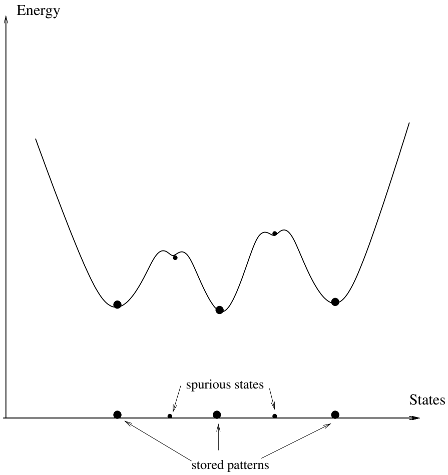
The table shows the probability of error for some values of p/N.

P_{error}	p/N
0.001	0.105
0.0036	0.138
0.01	0.185
0.05	0.37
0.1	0.61

Spurious states

- Therefore, for small enough p, the stored patterns become attractors of the dynamical system given by the synchronous updating rule.
- However, we also have other, so-called **spurious states.**
- Firstly, for each stored pattern $\vec{x}^k$, its negation $-\vec{x}^k$ is also an attractor.
- Secondly, any linear combination of an odd number of stored patterns give rise to the so-called **mixture states**, such as
$$\vec{x}^{\text{mix}} = \pm \text{sgn}(\pm \vec{x}^{k_1} \pm \vec{x}^{k_2} \pm \vec{x}^{k_3})$$
- Thirdly, for large p, we get local minima that are not correlated to any linear combination of stored patterns.
- If we start at a state close to any of these spurious attractors then we will converge to them. However, they will have a small basin of attraction.

Energy landscape



- Using a stochastic version of the Hopfield model one can eliminate or reduce the spurious states.

Strong patterns

- Suppose a pattern $\vec{x}$ has been stored $d \geq 1$ times in the network, we then call $\vec{x}$ a strong pattern if $d > 1$ and a simple pattern if $d = 1$.
- The notion of strong patterns has been recently introduced in Hopfield networks to model behavioural and cognitive prototypes, including attachment types and addictive types of behaviour, in human beings as patterns that are deeply or repeatedly learned.
- A number of mathematical properties of strong patterns and experiments with simulations indicate that they provide a suitable model for patterns that are deeply sculpted in the neural network in the brain.
- Their impact on simple patterns is overriding.
- $$P_{\text{error}} = \frac{1}{2}(1 - \text{erf}(d\sqrt{N/2p}))$$

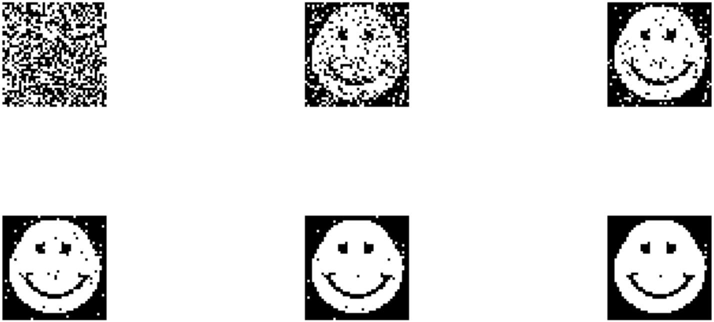
Experiment

- Consider a network of 48×48 pixels.
- We train the network with 50 copies of a happy smiley

- 30 copies of a sad smiley

- and single copies of 200 random faces.
- We then expose it to a random pattern to see what pattern is retrieved.

The strongest learned pattern wins!



References

- Devaney, R. L. *An Introduction to Chaotic Dynamical Systems.* Westview Press, 2003.
- Barnsley, M. F. *Fractals Everywhere.* Morgan Kaufmann Pub; 2 Sub edition, 1993.
- Flake, G. W. *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation.* The MIT Press, 1998.
- Hertz, J. A. and Krogh, A. S and Palmer, R. G. *Introduction To The Theory Of Neural Computation*, Westview Press, 1991.
- Edalat, A. and Mancinelli, F. *Strong Attractors of Hopfield Neural Networks to Model Attachment Types and Behavioural Patterns*, Proceedings of International Joint Conference on Neural Networks (IJCNN) 2013.

- Note that some of the pictures in the notes have been taken from the above.