

Physics of Complex Systems

https://github.com/Grufoony/Physics_Unibo

Contents

1	Introduction to Complex Systems	5
1.1	Emergence	5
1.2	Network's energy	7
1.3	Broken Stick Model	9
1.4	Rankings	10
1.5	Laplace problem	12
2	Basic models	13
2.1	Linear models	13
2.2	The neuron model	15
2.2.1	Hopf Bifurcation	16
3	Dynamical Systems	17
3.1	Malthus model	18
3.2	Velhurst model	18
3.3	Expanding map	19
3.4	Classification of chaotic systems	20
3.4.1	Topological transitivity	21
3.4.2	Rotation on a d-thorus	22
3.4.3	The Kronecker flow	24
3.4.4	Important map	25
3.4.5	Arnold's cat map	26
4	Ergodic theory	29
4.1	Birkhoff's ergodic theorem	31

Chapter 1

Introduction to Complex Systems

Complex systems are interdisciplinary: you have to communicate your results to experts from different science field. You need to provide an easy answer to a complex problem, no matter how difficult it's to get this answer.

1.1 Emergence

One of the main ideas in complex systems is emergence. Emergence means that the structure of the particles is simple and they are not so important like interactions between them.

An example of this is the Central limit theorem, which comes from mathematics.

If you have a random variable x_k , with average value $\langle x_k \rangle = 0$ and finite variance σ^2 , then the central limit theorem says that, if the variables are independent (and in physics this is usually a fair assumption) for every value of k , then the normalized sum

$$z_n = \frac{1}{\sqrt{N}} \sum_{k=0}^N x_k$$

then the distribution of this variable z_n is known, and it is gaussian, for a big enough value of N .

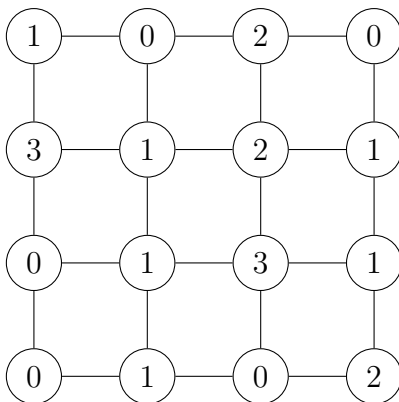
$$\rho(z_n) \sim \exp\left(-\frac{z_n^2}{2\sigma^2}\right)$$

Despite the fact that we, as humans, need a cause-effect relationship to describe a phenomenon, nature loves independent events, like DNA mutations. The gaussian describes the fluctuations of a system at equilibrium, not the complexity. You cannot extract work from fluctuations at equilibrium, otherwise you violate thermodynamics (and that's no good).

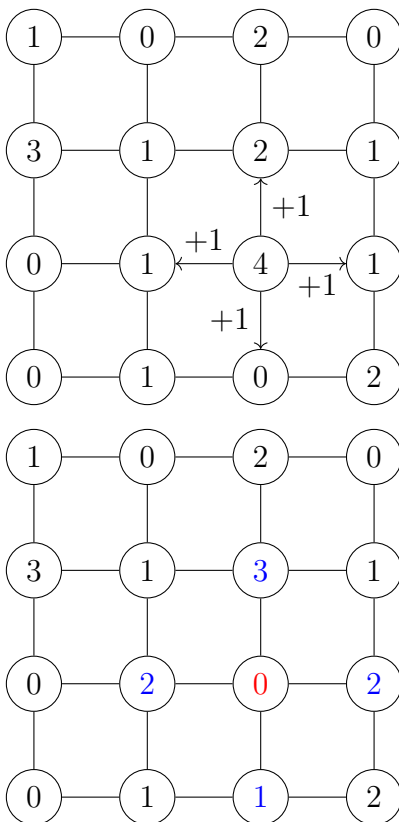
The gaussian function is not a physical function, because it implies non zero probabilities to events which are impossible. For example, if we take a particle in a basin of attraction of a potential, the non zero probability given by the gaussian fluctuations

allows the particle to jump out of the pit. But this violates the second law of thermodynamics.

A way to defy the gaussian properties is to allow a system to have memory, so to remove the independence of the variable.



One example of this is the sand pile model. You have a lattice, and each point is connected to its four neighbors. At this point one particle is put in a randomly chosen point. Each node can have four possible states, 0, 1, 2, 3, that is four possible numbers of particles. If a node reaches 4 particles, the 4 particles are distributed to the 4 neighbouring nodes.



For the nodes in the border, what happens is that 2 of the particles are redistributed and the other 2 are released in the environment, so they are lost.

So this system has memory, and this means that the states of all the nodes are not independent. This memory turns the gaussian distribution into a power law

$$p(n) \propto n^{-\alpha}$$

with $\alpha > 1$.

The decay of the power law is much slower than that of the gaussian, which means that the probability to have events in the extremes is significantly higher with the power law distribution. Memory is related to power laws, but vice versa is not guaranteed. We can also have power laws in physics, for example in the Ising model. In physics, power laws typically represent a phase transition for a dynamic system in a non-equilibrium state.

In the sand pile model, we notice a self-organized criticality: the system naturally goes into a critical state and does a phase transition.

Another example of a natural power law is the Kleiber law which relates the metabolic rate (amount of energy you need to survive) to the mass.

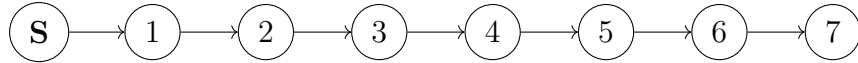
$$E(m) \propto m^{\frac{4}{3}} \quad (1.1)$$

We can also observe that heartbeat rate decreases with mass and lifetime increases.

1.2 Network's energy

Let's take a region of space containing a number N of nodes. This system can represent, for example, the hydraulic network of a city. For a node we define its "energy" flow φ (in the case of the hydraulic network, what is flowing between the nodes is water). How much "energy" do we need to insert in the system?

If we have to distribute something to all the nodes, the most basic way to do it is to connect one on one all the nodes, as to form a long chain.



If the flow out of the source S is ϕ , the flow after the first node is $\phi - \varphi$, and so on, and we expect that the flow arriving to the final node will be φ .

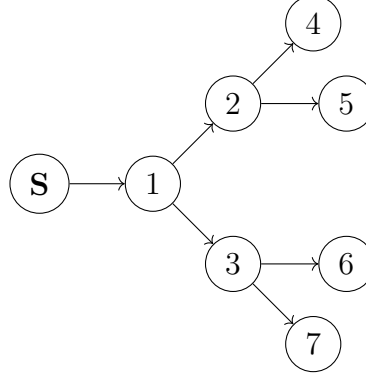
So the total energy is

$$E_T \approx l \sum_{k=0}^{N-1} (\phi - k\varphi) = l\varphi \sum_{k=0}^{N-1} k$$

$$E_T \propto l\varphi N^2$$

So the total energy that is required to provide for all the nodes, with this configuration, is proportional to the square of the total number of nodes. So with this configuration we have a network that is very easy to implement, but the network is very inefficient.

Another possible connection of the nodes is the following one:



So each node is linked to two more nodes. This is what is called a tree structure. The main problem with such a structure is that the path between the nodes is always the same, but it is impossible to design a city in such a way. The number of nodes is $N = 2^{m+1}$, where m is the number of levels. The relation for the flow in each level is

$$2\phi_k + \varphi = \phi_{k-1}$$

where we assume again that the flow in the final level is φ . ϕ_k then is

$$\phi_k = (2^{m+1-k} - 1)\varphi$$

Then:

$$E_T = d\varphi \sum_{k=1}^m 2^k (2^{m+1-k} - 1) \approx d\varphi m 2^{m+1}$$

$$E_T = d\varphi N \log_2 N$$

This configuration is much more efficient than the previous alternative, although it is impossible to put to practice.

So we need to find a solution that is in the middle of these two.

Another possible structure consists of connecting all the nodes to the source, making a star network: also this model is of course impossible to put in practice. However, in this case the total energy required is proportional to NL , where L is the scale length of the system.

$$E_T \propto NL$$

The proportionality to L is due to the fact that the average length of the links is proportional to the size of the space.

Since

$$N = \left(\frac{L}{l}\right)^D$$

where D is the system dimension (e.g. 1D, 2D...). We have that

$$E_T \propto N\bar{L} \approx N^{1+1/D} = V^{\frac{D+1}{D}}$$

We notice that:

- for a 2D system, like a city, $E \propto V^{\frac{3}{2}}$
- for a 3D system, like biological systems, $E \propto V^{\frac{4}{3}}$

In any case the exponent is greater than one so, soon or later, the system will collapse. The problem is also logistic, because in a real transportation line there is also waste, that has to be disposed of.

If we consider our system to be an animal's body, this would mean that in order to survive the amount of blood in its body must grow with the size by an amount of $4/3$. This would mean that very big animals can't survive, but they do, and the solution is in Kleiber's law.

As animals get bigger, their demand of energy goes down, so their metabolism slows down.

Once we have a model (which hopefully is a good one), the model is deterministic, so if we know the initial conditions of the system, we know exactly what state the system will have at any future time.

The problem with this is that we don't know the initial conditions with absolute certainty. So at this point the question, that will be answered later, is:

How much error in the initial conditions is acceptable before that the previsions of our model start to diverge from the real evolution in an unacceptable way?

1.3 Broken Stick Model

Let's consider the portion $[0, 1]$ of the axis, which represents a segment (a stick) of length 1. Suppose that we extract randomly a point x_1 on that segment, and we cut it in correspondence of that point, thus obtaining the portion $[0, x_1]$ of the axis.

By repeating this process many times, we obtain a system with memory, because of course the length of the segment at a certain iteration depends on all the previous iterations.

This is called the broken stick model.

For this system one expects to have a power law distribution, because if we rescale the variable x , the distribution must not change.

$$p(x) \sim \frac{1}{x^a}$$

$$y = \lambda x \longrightarrow p(y) = \frac{1}{(\lambda x)^a} \sim \frac{1}{x^a}$$

Since we have:

$$\langle x_k \rangle = \frac{\langle x_{k-1} \rangle}{2}$$

$$\rho_N(2x) = \frac{1}{2} \rho_{N-1}(x)$$

Which means that, as N goes to infinity we get

$$\rho(2x) = \frac{1}{2} \rho(x)$$

$$\rho(x) = \frac{1}{x}$$

Let's now try to implement the system without memory. So we have N independent variables x_k uniformly distributed, and for each variable we define its distance from the previous one, Δx .

The probability to find Δx is the probability of not finding x in any segment, so

$$p(\Delta x) \approx \left(1 - \frac{\Delta x}{L}\right)$$

We then take a new variable $y = N\Delta x$ and we obtain the probability distribution

$$p = \exp(-y/L)$$

as N goes to infinity, which of course is an exponential law.

Suppose that the whole stick is a state, and we want to distribute the population inside of it. If we divide the stick uniformly in portions and divide the populations in this group, we obtain an exponential law, as we have just seen.

Another way, which contains memory, consists of creating a city and letting it grow, and only then introducing a second one, and repeating the process until all the space is occupied.

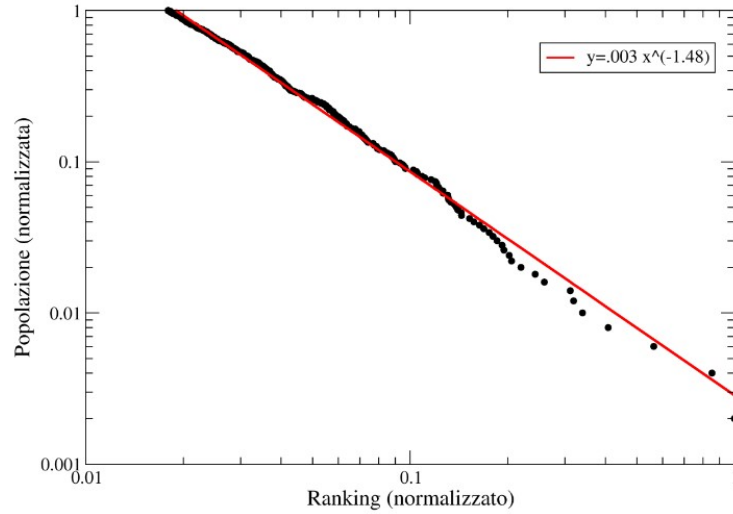
1.4 Rankings

A ranking is an ordered subset of measures of the same quantity, between which exists an order relation. Rankings allow us to obtain complex informations with the use of

very little data, since the information is contained in the order of the values.

An example of the use of rankings can be found in web pages, where the search engine shows to the user the results of the search, and the results are ordered depending on their relevance, so to allow them to find the desired information as quickly as possible. In this example, the quantity is the relevance of a result to the search.

Another example is the population of cities. The graph below shows the population of the 300 biggest cities in Europe. The graph is in log-log scale, and since it shows a straight line, it means that the distribution is a power law distribution.



This can be explained considering the fact that, historically, the population used to migrate towards the biggest cities of the country, because there were more job opportunities. Thus one can say that big cities have a preferential attachment.

Let x be a random variable and $\{x_1, x_2, \dots, x_N\}$ a sample of experimental observations of the variable, ordered in such a way that $x_1 \geq x_2 \geq \dots \geq x_N$.

Then, the ranking distribution is created by assigning to the position j in the order the corresponding value in the sample, which is given by the map $x_j = f(j)$. Also, usually the elements are normalized like $y_j = x_j/x_1$.

By definition, j/N is the frequency of the event $\{x \mid x \geq x_j\}$, that is, the probability that the variable x has a value larger than x_j for larger values of j .

The cumulative distribution is defined as $F(x) = P\{u \mid u \leq x_j\}$, that is, the probability that the value u is smaller than x . Of course, this probability is equal to the integral of the distribution in the range $[-\infty, x]$.

The relation between the frequency j/N and the cumulative distribution is given by:

$$F(x_j) = 1 - \frac{j}{N} = 1 - \frac{J(x_j)}{N}$$

where $J(x_j)$ is the inverse of the ranking, so it's the function that returns the position of a certain value of x in the sample's order.

Since the cumulative distribution is the integral of the probability distribution, the latter can be obtained by differentiating the first:

$$p(x) = \frac{dF}{dx} \longrightarrow p(x) = -\frac{1}{N} \frac{dJ}{dx}$$

1.5 Laplace problem

Let's suppose that we are tossing a coin, and the probability of getting head is p and the probability of getting tails is $1 - p$. The event E is x_{n+1} outcome, and H_n is $\{x_1, x_2, \dots\}$, so the first is the future and the second is the past.

Now we ask ourselves what is the probability that the next result of the toss is going to be heads:

$$p(x_{n+1} = H | H_n)$$

The most intuitive way to calculate this probability would be to use the frequentistic definition, so to count how many times we have gotten heads over the entire number of tries (we are considering the number of tries to be very big of course)

$$p = \frac{n_H}{N}$$

Unfortunately, this is not the right way to do it.

The probability of getting a given sequence of heads and tails is

$$p(\{x_k\}) = p^J (1 - p)^{n-J}$$

where J is the number of times that we get heads.

Now if we integrate we get

$$\int_0^1 p^J (1 - p)^{n-J} dp = \frac{J!(n - J)!}{(n + 1)!}$$

so

$$p(x_{n+1} = H | \{\}) = \frac{(n + 1)!}{J!(n - J)!} p^J (1 - p)^{n-J}$$

which is a binomial distribution.

Now we calculate the average value of the probability

$$\int_0^1 p \frac{(n + 1)!}{J!(n - J)!} p^J (1 - p)^{n-J} dp = \frac{J + 1}{n + 2}$$

and we see that it isn't J/n , as one would have expected.

Chapter 2

Basic models

2.1 Linear models

The most simple model possible is the linear model

$$\dot{x} = Ax$$

Albeit this model is easy, there are a few complications that have to be taken into account:

- When the number of dimensions increases it becomes complicated.
- If we consider the determinant

$$\det(\lambda I - A) = 0$$

we obtain a polynomial equation, that can be very hard to solve.

- To further complicate things, A might not be known, but we could have an ensemble of matrices.
- The point $x = 0$ is always a critical point, so we can always linearize the system around this point, but if A has critical points then we get those as well.

By solving the determinant equation, we get the eigenvalues λ , whose study can tell a lot about the behaviour of the solutions.

If $\operatorname{Re}\lambda \leq 0$, this means that the exponential term of the solution shrinks, so it converges.

Robustness means that if we perturb the system, the solutions don't change too much. A model must be robust, otherwise it can fit any kind of data simply by slightly changing the parameters.

The formal way to write the solution of such a linear system is

$$x(t) = x_0 \exp(At)$$

For coefficients λ^* such that $Re\lambda^* > 0$, we have that

$$|\delta x| \approx |e^{\lambda^* t}| |\delta x_0|$$

This is very typical, and from this rises the chaos theory. In this case, even a small error in the initial condition will increase the error in the model exponentially fast.

We get another interesting model by adding some noise to the linear model

$$\dot{x} = Ax + \xi(t)$$

We consider a special solution of the form

$$x = e^{At}y$$

and we substitute in the equation

$$\begin{aligned}\dot{x} &= A \exp(At)y + \xi(t) = A \exp(At)y + \exp(At)\dot{y} \\ \dot{(y)} &= \exp(-At)\xi(t)\end{aligned}$$

So the complete solution of the problem is

$$x(t) = \exp(At)x_0 + \int_0^t \exp(A(t-s))\xi(s)ds \quad (2.1)$$

If the system is stable (negative real part of lambda) and we use a periodic forcing, the solution is still a periodic function with different period.

Now we want to consider the case of a perturbed matrix:

$$\dot{x} = (A + \varepsilon B)x$$

with $\varepsilon \ll 1$.

The solution is

$$x(t) = \exp((A + \varepsilon B)t)x_0$$

and to understand the system's sensitivity we calculate the derivative for small perturbations

$$\frac{d}{d\varepsilon} \exp((A + \varepsilon B)t)$$

but the problem is that usually the two matrices are not commutative.

If we take the special solution $x = \exp(At)x_0$ and we substitute we get

$$\dot{y} = \varepsilon \exp(-At)B \exp(At)y$$

and, if A and B do not commute, this equation is very difficult to solve. An approximate solution is

$$y(t) = y_0 + \varepsilon \int_0^t \exp(-As) B \exp(As) y_0 ds + O(\varepsilon^2)$$

and the solution for x is

$$x(t) = \exp(At)x_0 + \varepsilon \int_0^t \exp(A(t-s)) B \exp(As)x_0 ds + O(\varepsilon^2) \quad (2.2)$$

Now, the sensitivity of the system is

$$\frac{dx}{d\varepsilon} = \int_0^t \exp(A(t-s)) B \exp(As)x_0 ds$$

In the particular case where the matrix is diagonal, the sensitivity becomes

$$\frac{dx}{d\varepsilon} = \int_0^t \exp(\lambda_i(t-s)) B_{ij} \exp(\lambda_j s) x_0 ds$$

2.2 The neuron model

In this model we have two variables, x and w , and two differential equations to describe their dynamics

$$\begin{aligned} \dot{x} &= x - \frac{x^3}{3} - w + I(t) \\ \dot{w} &= \frac{1}{\mu}(x + a - bw) \end{aligned}$$

This system is of course not linear, because we have a cubic term in the dynamics of x . We require μ to be big, so that w is the slow variable, whereas x is the fast variable. The fact that the two variables have different growth speeds is quite important, because it allows us to study them separately, and this reduces the complexity of the system. The difference in the speed of the variables makes the system stiff, which means that it's very hard to integrate numerically.

The key to studying this system is in the nullcline. We find the two nullclines by putting the two derivatives equal to zero, so one is a straight line and the other one is a cubic.

The two nullclines intersect, and the point where they intersect is the equilibrium point of the system. Furthermore, the cubic nullcline has two critical points. Since x is the fast variable, the point follows the cubic nullcline in its dynamics. Once it reaches one of the two critical points, during its motion, it jumps to the other branch of the nullcline. So in the end one gets a periodic motion.

Suppose now that $I(t)$ is of the form

$$I(t) = I_0 + \Delta I \delta(t)$$

so we give an impulse to the point.

What happens now is that, if ΔI is small, the point oscillates briefly around the critical point and falls back into it, but if the impulse is big, the point escapes and makes a very big oscillation.

In this system we have an Hopf bifurcation.

2.2.1 Hopf Bifurcation

Sometimes the orbits of a dynamical system converge not in a single point but in limit circle near the stable point. This is called *Hopf Bifurcation*. That type of phenomena is recurrent in nature: one example may be the Ising's model in which the system passes from a one stable point state to a two stable points state with a phase transition. In that specific case we observe a bifurcation of the system's free energy.

We have to see the Hopf bifurcation as a new dynamical state that has a periodic orbit as a solution. However, it's important to observe that a Hopf bifurcation is not an equilibrium state for the system.

For example, consider a circular network of identical neurons characterized by a function \dot{x}_k . The dynamic is described by a simple law

$$\dot{x}_k = F(x_k) + \epsilon x_{k-1}$$

in which the function $F()$ is taken from the Fitzhugh-Nagumo model and the other term represents the system dynamic.

If x_P is an equilibrium point for all neurons then we can verify that exists an $\epsilon = \epsilon_C$ that produces a non-trivial solution. In particular, with that value of ϵ we can create a stationary periodic state with a periodic signal (moving wave): that's a self-solution for the system. Obviously, this example do not conserve the total energy of the system.

Chapter 3

Dynamical Systems

A dynamical system is given by a *phase space* $M \subseteq \mathbb{R}^d$ endowed with a collection of maps $\phi^t : M \rightarrow M$, where $t \in \mathbb{G} = \{\mathbb{R}, \mathbb{C}, \mathbb{Z}, \mathbb{N}\}$, with the so-called “group property”:

$$\phi^{t+s} = \phi^t \circ \phi^s$$

The maps ϕ^t are usually called *flows*.

This formalization works for systems whose equations don’t depend explicitly on time.

A typical example of a dynamical system is that of the solutions of a differential equation. So if we have a Cauchy problem

$$\begin{cases} \dot{x} = f(x) \\ x(0) = x_0 \end{cases}$$

where the flow is the function that takes the initial condition and returns the correct solution for the differential equation. So if this ODE has a global solution $x(t)$ that satisfies the initial condition, then

$$\phi^t(x_0) = x(t)$$

We must also assume the uniqueness of the solution, that is, for one initial condition we only have one solution.

Non-autonomous ODEs, where the equation field depends on time

$$\dot{x} = f(x, t)$$

can be converted into autonomous ODEs in higher dimensions, going back to the previous simpler case.

We consider $\overline{M} = M \times \mathbb{R}$ with $y = (x, t) \in \overline{M}$, and we define

$$F(x, t) = (f(x, t), 1)$$

with $f : M \times \mathbb{R} \rightarrow \mathbb{R}^d$ and $F : M \times \mathbb{R} = \overline{M} \rightarrow \mathbb{R}^{d+1}$. At this point we can write the new differential equation:

$$\dot{y} = F(y)$$

with initial condition $y(0) = (x_0, 0)$. For this to be useful, we must show that knowing the solution to this new equation implies knowing the solution for the original one.

Say $y(t) = (x(s), t(s))$ is a solution for the new equation

$$\begin{pmatrix} \dot{x}(s) \\ \dot{t}(s) \end{pmatrix} = F \begin{pmatrix} x(s) \\ t(s) \end{pmatrix} = \begin{pmatrix} f(x(s), t(s)) \\ 1 \end{pmatrix} =$$

$$\dot{x} = f(x, s) \quad \dot{t} = 1$$

with initial conditions $x(0) = x_0$ and $t(0) = 0$. For what concerns t , it's the identity function, so $t(s) = s$, which means that I can invert t and s , so going back to the first equation we have

$$\dot{x} = f(x, t)$$

which was our original equation. This means that a solution for the non-autonomous equation is also a solution for the new autonomous one.

So, if we have a theory that treats non-autonomous systems, that theory must also be able to treat autonomous systems.

If the t variable is a natural number, we call it n

$$\phi^n = T^n$$

3.1 Malthus model

Let's take the expression

$$x_{n+1} = \alpha x_n$$

where x_n represents the population in year n . For $\alpha > 1$, we have population growth and for $\alpha < 1$ the population decreases.

If at time $t = 0$ the population is x_0 , at a given time we have

$$x_n = T_n(x_0) = \alpha^n x_0$$

3.2 Velhurst model

This model takes into account the fact that the higher the population is, the more resources are needed, so there should be a limit that can't be surpassed, because the environment wouldn't be able to sustain it

$$x_{n+1} = \alpha x_n (A - x_n)$$

Now we change variable, defining $y = x/A$, and we get

$$\frac{x_{n+1}}{A} = A\alpha \frac{x_n}{A} \left(1 - \frac{x_n}{A}\right)$$

$$y_{n+1} = ry_n(1 - y_n)$$

which is the logistic map:

$$T(x) = rx(1 - x) \quad (3.1)$$

We take $0 < x < 1$ and $1 < r < 4$. The graph is a parabola upside down, with the maximum at $(\frac{x}{2}, \frac{r}{4})$. It may happen that $T(x)^n = x$: in this case the orbit $O(x) = \{\Phi^t(x)\}_{x \in \mathbb{G}}$ is called *periodic orbit*. We also observe that getting closer to an orbit is not the same as getting closer to a limit.

3.3 Expanding map

Now let's see an example of a chaotic map. First of all, we define a map T as *expanding* if $|T'(x)| > 1 \forall x \in \mathcal{M}$. As well, we also give this definition for a generic point:

Definition 1. We define a *periodic point* of the map $x_0 \in \mathcal{M}$ as:

- attractive if $|T'(x_0)| > 1$
- repelling if $|T'(x_0)| < 1$
- neutral if $|T'(x_0)| = 0$

For example, let's take the map $T : [0, 1] \rightarrow [0, 1]$ such as $T(x) = 2x \bmod 1$. This map has two fixed points (intersection with the bisectrice): $p_0 = (0, 0)$ and $p_1 = (1, 1)$. We notice immediately that those points are both *repelling* points. Another thing we can see with this particular map is that if we start on a slightly different point from the previous one, for example $x \rightarrow x + \epsilon$, the difference between the two orbit grows exponentially fast:

$$T^n(x) = 2^n(x - k(x)) - k(T(x))$$

where

$$k(x) = \begin{cases} 0 & 0 \leq x \leq \frac{1}{2} \\ 1 & \frac{1}{2} < x \leq 1 \end{cases}$$

and in conclusion we have that

$$T(x + \epsilon) = 2^n(x + \epsilon) \bmod 1$$

so the difference grows $\propto 2^n \epsilon$. One last comment to make is that it's true that the two orbit exponentially diverges one from the other but the phase space is limited: from this point of view their distance in phase space will oscillate through time.

Proposition 1. If $T : [0, 1] \rightarrow [0, 1]$ is an expanding map such as $|T'(x)| > 1 \forall x \in [0, 1]$ then any point p of T^j ($j \in \mathbb{Z}$) is repelling.

Proof. If we take the derivative of T^j we get

$$(T^j)'(x) = T'(T^{j-1}(x)) \times T'(T^{j-2}(x)) \times \dots \times T'(x) = \prod_{k=0}^{j-1} T'(T^k(x))$$

and if now we calculate the module of this derivative we get

$$\left| (T^j)'(x) \right| = \prod_{k=0}^{j-1} |T'(T^k(x))| \gg 1$$

where, since the map is expanding, each term of the product is larger then one. \square

Proposition 2. *If two points p and q are part of the same periodic orbit then they are both of the same type.*

Proof. Let's consider an orbit of period j with $q = T^k(p)$, $0 < k \leq j-1$.

$$(T^j)'(p) = T'(p_0) \dots T'(p_{j-1})$$

$$(T^j)'(q) = (T^j)'(p_k) T'(p_k) \dots T'(p_{k-1}) \dots T'(p_{k+j-1})$$

Real numbers always commute, so it's obvious that

$$(T^j)'(q) = (T^j)'(p)$$

\square

We have just shown that the property of a point (local) is also a property of an orbit (global).

3.4 Classification of chaotic systems

Let π^1 be a one-dimensional torus (a circle). We define the transformations

$$R_\alpha : \pi^1 \rightarrow \pi^1, R_\alpha(x) = (x + \alpha) \bmod 1$$

which are called rotations by an “angle” α .

Now, if $\alpha = p/q$, with $\gcd(p, q) = 1$, then:

- all orbits are periodic with fundamental period q .
- if α is not rational, then all orbits are dense.

Proof. • We consider that $R_\alpha^n(x) = x$ is equivalent to $(x + n\alpha) \bmod 1 = x$, so

$$\left(x + n\frac{p}{q}\right) \bmod 1 = x$$

which means that,

$$n\frac{p}{q} = k \longrightarrow n = \frac{kq}{p}$$

and for n to be an integer, p needs to be simplified, but it can't be simplified with q , so it simplifies with k , which is then equal to $k = mp$. So in the end we get

$$n = mq$$

- In order to prove that all orbits are dense, we prove that they are ε -dense for all ε .

Since α is not rational, the orbit must be infinite, because it is not periodic. Then, we can say that there is at least one accumulation point in the orbit.

Now, call $j = n - m$, then we can say that $\{R_\alpha^{k_j}\}_k$ is ε -dense in π^1 , which means that the orbit is dense.

□

3.4.1 Topological transitivity

Definition 2. A dynamical system (M, ϕ^t) is called topologically transitive (or just transitive) if exists $x \in M$ such that the future orbit

$$O^t(x) = \{\phi^t(x)\}$$

is dense.

So a transitive system is a system that has at least one dense orbit. Going in that direction, if all the orbits are dense we say that the system is minimal.

Definition 3. A system is called minimal if all forward orbits $O^t(x)$ are dense.

Another very important definition is that of an invariant subset:

Definition 4. A subset $A \subseteq M$ is called invariant if $\phi^{-t}(A) = A$ for every t .

Note that if ϕ^t is invertible, this statement is equivalent to saying that $\phi^t(A) = A$ for every t .

Finding invariant subsets is important because we know that if we have an initial condition in that region, the orbit is going to stay in there. Of course, this doesn't give us the trajectory, but it helps us by restricting the phase space.

A function $f : M \rightarrow R$ (observable) is called invariant if $f \circ \phi^t = f$ for every $t \geq 0$.

This is what is called in mechanics a prime integral of motion.

Suppose $f = 1_a$, where $1_a(x) = 1$ if $x \in A$ and $1_a(x) = 0$ if $x \notin A$. Then we have that

$$1_a \circ \phi^t = 1_a$$

but also

$$1_a \circ \phi^t = 1_{\phi^{-t}(A)}$$

So combining the two we get that

$$\phi^{-t}(A) = A$$

Proposition 3. *If a dynamical system is topologically transitive, there exists no continuous invariant observables other than constant functions.*

Proof. Say that $O^t(x_0)$ is dense. Now, taking any $y \in M$, we can say that exists a sequence $x_j = \phi^{t_j}(x_0)$ such that $x_j \rightarrow y$.

Now, suppose that a function f is continuous, then

$$f(x_j) \rightarrow f(y)$$

and suppose that f is invariant, so that

$$f(x_j) = f(\phi^{t_j}(x_0)) = f(x_0) = C$$

Since we have taken f to be continuous, $f(x_j)$ must converge to $f(y)$, but since it is constant and equal to C , it can only converge to C , which means that

$$f(y) = C$$

which is a constant function. □

3.4.2 Rotation on a d-thorus

Now we try to generalize rotations of the circle to translations of a d-thorus.

We define the translation on a d-thorus T_γ as:

$$T_\gamma : \pi^d \rightarrow \pi^d, T_\gamma(x) = (x_1 + \gamma_1, x_2 + \gamma_2, \dots) \bmod 1 = (x + \gamma) \bmod 1$$

Definition 5. *A collection of real numbers $\{\lambda_1, \lambda_2, \dots\}$ is said to be rationally independent if none of them can be written as a combination of the others with rational coefficients.*

Now that we have defined the concept of rational dependence, we can use it to introduce a characteristic of the translations on d-thori in the case where the vector components are rationally independent:

Proposition 4. *Given a translation on a d -torus T_γ , it is minimal if and only if $\{1, \gamma_1, \gamma_2, \dots\}$ are rationally independent.*

In this case, topological transitivity (so the fact that the system has one dense orbit) is equivalent to minimality (so all the system's orbits are dense). This is due to the fact that translations commute. In fact, if we define a translation of the initial point x_0

$$T_{x-x_0}(x_0) = x$$

then we see that translating the orbit of T_γ starting from x_0 gives an orbit which is identical to the orbit starting from x :

$$O^t(x_0) = O(x)$$

$$T_{x-x_0}O^t(x_0) = T_{x-x_0}(T_\gamma^k(x_0)) = T^k(T_{x-x_0}(x_0)) = O(x)$$

which means that the translated orbit is dense if and only if the initial one is dense, so translating a dense set we get another dense set.

So we only need to show topological transitivity, instead of minimality. Let's show that topological transitivity implies rational independence: By contradiction, suppose that $1, \gamma_1, \gamma_2, \dots$ are not rationally independent, which means that exists a set k_1, k_2, \dots , where k_i are not all zero, such that

$$\sum_i k_i \gamma_i = 0$$

We then define the function $f(x) = e^{2\pi i \langle k, x \rangle}$, which is continuous and non constant, because not all k_i are zero (so you have at least one x_i in the exponent). Now we just need to show that f is invariant. To do this we calculate $f(T_\gamma(x))$:

$$\begin{aligned} f(T_\gamma(x)) &= \exp(2\pi i \langle k, x + \gamma \rangle) = \exp\left(2\pi i \sum_i k_i x_i\right) \exp\left(2\pi i \sum_i k_i \gamma_i\right) \\ &= \exp\left(2\pi i \sum_i k_i x_i\right) = f(x) \end{aligned}$$

and this holds for every x in the torus.

Let's take a system on a ring, and the dynamics represents rotations on this ring. We can imagine this as a cave tube filled with fluid (marmellade). In this case, the motion of the fluid will be laminar.

In discrete time, we take a surface with codimension 1, which intersects the trajectory in a point x , and we want to find after how many iterations the trajectory intersects the surface again, so in other words, how long it's going to take for the trajectory to

intersect the surface again.

A Poincarre section of a continuous dynamical system is a codimension 1 submanifold such that all trajectories cross it transversally (more or less).

In this case, the so called Poincarre map, also known as first return map, is defined as follows:

$$T_{M_0} : M_0 \rightarrow M_0, \quad T_M(x) = \phi^{t_{M_0}(x)}(x) = \min\{t > 0 | \phi^t(x) \in M_0\}$$

where t_{M_0} is the return time to M_0 .

3.4.3 The Kronecker flow

Continuous time dynamical system, the Kronecker flow.

$$\phi^t : \pi^d \rightarrow \pi^d$$

given by the solutions of the ODE

$$\dot{x} = \omega$$

In this case it's very easy to find the flux, which is simply

$$\phi^t(x) = (x + \omega t) \bmod 1$$

In general terms, suppose that $\omega = (\omega_1, \omega_2, \dots, \omega_d)$ with $\omega_d > 0$. Then the bottom side of the torus M_0 is

$$M_0 = \{(x_1, x_2, \dots, x_{d-1}, 0) \in \pi_d\} = \pi_{d-1}$$

and we know that all trajectories must cross this subspace.

The first return map is

$$T_{M_0} : M_0 \rightarrow M_0, \quad T_{M_0}(x) = \phi^{t_{M_0}(x)}(x) = x + \frac{\omega}{\omega_d} \bmod 1$$

This is the same as

$$T_\gamma : \pi^{d-1} \rightarrow \pi^{d-1}$$

with $\gamma = (\omega_1/\omega_d, \omega_2/\omega_d, \dots)$.

Now we ask ourselves, when are the orbits of T_γ dense? The answer is that they are when $\gamma_1, \gamma_2, \dots, \gamma_{d-1}, 1$ are rationally independent. But this condition is met if and only if $\omega_1, \dots, \omega_d$ are rationally independent.

This means that the trajectories of ϕ^t are dense if and only if $\omega_1, \dots, \omega_d$ are rationally independent.

3.4.4 Important map

$$T_2(x) = T(x) = 2x \bmod 1$$

We can think of this map as either a map that goes from $[0, 1]$ to itself, or as a map on a torus. Alternatively it can be written as

$$T(x) = 2x, \quad 0 \leq x \leq \frac{1}{2}$$

$$T(x) = 2x - 1, \quad \frac{1}{2} \leq x \leq 1$$

If we consider it as a map on a torus, it is quite easy to see that this map is continuous. As we have seen, this map is chaotic. To understand, if we consider two points x_1 and x_2 , positioned at a distance ε , their distance after one iteration becomes 4ε .

We understand that after a number of iterations these two points diverge pretty quickly, no matter how small we take ε to be.

We write x as $x = 0.d_1d_2d_3$, where d_i are the digits. It can be written as

$$x = \sum_{i=1}^{\infty} 2^{-i} d_i$$

We define the map $T(x)$ as

$$T(x) = T(0.d_1d_2d_3) = 0.d_2d_3$$

Definition 6. A topological dynamical system (a system for which we care about topological properties) is said to be topologically mixing if $\forall U, V$ non-empty open sets, $\exists N = N(U, V) \in \mathbb{N}$ such that

$$\phi^t(U) \cap V \neq \emptyset$$

for all t .

Proposition 5. (M, ϕ^t) is topologically transitive if and only if $\forall U, V$ non-empty open sets $\exists N = N(U, V)$ so that $\phi^N(U) \cap V \neq \emptyset$.

A corollary of this proposition is that topological mixing implies topologically transitive.

We want to show that this map is topologically mixing. We know that a set, which is the arc connecting two points, expands in length after each iteration, and after a certain point it will fill the whole circle, and at that point it will keep being the whole circle, and in that case it's obvious that the intersection is not null. So we just need to prove that after a certain number of iteration the expanded set will fill the whole circle. The function after n iterations is

$$T^n(x) = 2^n(x) \bmod 1$$

For any n

$$T^n\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right) = [0, 1] = \pi^1$$

Since U is open, exists an n with $0 < k < 2^n - 1$ such that $\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right] \subset U$

The number of periodic points of period n of this map are $2^n - 1$.

3.4.5 Arnold's cat map

$$T : \pi^2 \rightarrow \pi^2, T(x) = T(x_1, x_2) = A(x_1, x_2) \bmod 1$$

where A is the matrix

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

If we look how the points $(1, 0)$ and $(0, 1)$ transform, we see that they go to $2, 1$ and $(1, 1)$.

This map preserves the Lebesgue measure μ

$$\mu(T^{-1}B) = \mu(B)$$

If the system is bijective, this is the same as saying

$$\mu(TB) = \mu(B)$$

so we can go forward in time. We know that the measure is preserved because $\det A = 1$. We study the eigenvalues and eigenvectors of this matrix. These are $\lambda_1 = \lambda = \frac{3+\sqrt{5}}{2}$ and $\lambda_2 = \lambda^{-1} = \frac{3-\sqrt{5}}{2}$, with corresponding eigenvectors $(1, \frac{\sqrt{5}-1}{2})$ and $(1, -\frac{\sqrt{5}+1}{2})$. The two eigenvectors represent the directions of maximum expansion and of maximum contraction.

If we take a point on the expanding eigenvector, its orbit diverges from the origin, whereas for a point on the contracting eigenvector the orbit is attracted to the origin. If we take a point which is a linear combination of the two directions, after each iteration the point gets closer and closer to the expanding eigenvector, because the component along the contracting eigenvector gets smaller after each iteration, and eventually it will disappear.

Proposition 6. *The cat map T_A is topologically mixing.*

This map is an example of an "hyperbolic dynamical system". Let's call

$$W^u(x) = \{y = (y_1, y_2) \in \pi^2 \mid y_2 - x_2 = \frac{\sqrt{5}-1}{2}(y_1 - x_1) \bmod 1\}$$

where x are constants, an unstable manifold. Then we can define the stable manifold as

$$W^s(x) = \{y = (y_1, y_2) \in \pi^2 \mid y_2 - x_2 = -\frac{\sqrt{5}+1}{2}(y_1 - x_1) \bmod 1\}$$

Why do we expect systems like this to have a lot of periodic orbits? Because of the stretching and folding mechanism. To see this, let's consider an hyperbolic rectangle, which is made up of pieces of stable manifolds in some points and unstable manifolds in other points. Now, we apply our dynamics to it, so the unstable manifolds will stretch and the stable ones will contract.

So if we have the rectangle R and the transformed rectangle $R' = T(R)$, R' will cross R at some point, with the condition that we can't have unstable manifolds crossing other unstable manifolds. Once we have found the intersection of the original rectangle and the transformed rectangle, we have found a fixed point of the transformation, hence we have found a periodic point, which implies the existence of a periodic orbit.

Chapter 4

Ergodic theory

Ergodic theory is the study of dynamical systems from the point of view of measure theory, hence of probability.

One first measure comes from analytical mechanics. We consider a measure space M , and in this space we have an energetic hypersurface. Then we define the Liouville measure as

$$d\mu = \frac{d\sigma}{|\nabla H|}$$

where $d\sigma$ is the measure of an infinitesimal portion of the hypersurface and ∇H is the gradient of the hamiltonian of the system.

We can derive the expression for the Liouville measure by considering the hypersurface with a bit of thickness in the direction of the gradient, forming in this way a volume A , and of course the system's dynamics is going to conserve the Lebesgue measure of A

$$Leb(A^t) = Leb(\phi^t(A)) = Leb(A)$$

If we then move in the direction of the gradient by a quantity h , we increase the energy by a quantity

$$\Delta H = |\nabla H|h$$

Then the conservation of the Lebesgue measure implies that

$$\sigma(B^t)h^t = \sigma(B)h$$

so we are not only on the same energy surface, we are in the same energy shell. We can write this as

$$\frac{\sigma(B^t)\Delta H}{|\nabla H_f|} = \frac{\sigma(B)\Delta H}{|\nabla H_i|}$$

and we can simplify the delta, so

$$\frac{\sigma(B^t)}{|\nabla H_f|} = \frac{\sigma(B)}{|\nabla H_i|}$$

and we have thus obtained the expression for the Liouville measure.

In ergodic theory we consider the space, along with the measure and the systems flux, (M, μ, ϕ^t) , or in discrete time (M, μ, T) .

The first thing that we could ask ourselves is the probability that $\phi^t \in A$. So now we don't consider a single trajectory but all the possible trajectories, and want to find the probability that they are contained in a certain portion of space. This probability is equal to

$$P(\phi^t \in A) = \mu(\phi^{-t}(A)) = \mu(\{x \in M \mid \phi^t(x) \in A\})$$

At this point we can define the concept of a measure preserved by the dynamics

Definition 7. *A measure is preserved by the dynamics if*

$$\mu(\phi^{-t}(A)) = \mu(A)$$

where A must be measurable, so it must belong to the σ -algebra of the space.

From now on we restrict to measures that preserve the dynamics, which are also called invariant measures.

An example of a measure preserving dynamical system is the doubling map. To see this, if we take the counter image of a set A , this is equal to the union of two sets A_1 and A_2 , which are measurable. In this case it is important to consider the counter image of the set, because the image wouldn't preserve the set

$$\mu(TA) \neq \mu(A)$$

When the two are equal, we have forward preservation.

An observable is any function of the phase space, $f : M \rightarrow R$, that represents any quantity of the system that can be observed and measured, thus giving us information about the system. An observable could be the temperature in a room, the moment of a particle or anything really.

Definition 8. *An observable $f : M \rightarrow R$ is said to be dynamics invariant (in ergodic theory) if*

$$f \circ \phi^t(x) = f(x)$$

We notice that this is equivalent to the definition that we had before, but in the study of dynamical systems we asked for the relation to hold for every value of x , but we can't do this in ergodic theory, because not all the points in the space are seen by the measure. In ergodic theory we say that a relation holds if it holds almost everywhere, that is, if it doesn't hold only on subsets of the space that have zero measure.

We also define the concept of an invariant set.

Definition 9. A set A is invariant if

$$\phi^t(A) = A \text{ mod } \mu$$

We can also write it as

$$\mu(\phi^{-t}A \Delta A) = 0$$

where

$$A \Delta B = (A \cup B) \setminus (A \cap B)$$

We then define push-forward measures:

Definition 10. We define a push-forward measure as

$$\phi^t * \mu(A) = \mu(\phi^{-t}(A))$$

The measure contains the counter image because we are considering the points at an initial time and this allows us to require that they will be in the set at forward times. We can say, in probability terms, that the probability of being in A at time t is the same as the probability of being in A at previous times.

Now we can note that a measure μ is preserved by the dynamics if and only if

$$\phi^t * \mu = \mu$$

Now, if $\phi^t * \mu \rightarrow \nu$, then ν is called an equilibrium measure.

4.1 Birkhoff's ergodic theorem

Given an observable $f : M \rightarrow \mathbb{R}$ we define the Birkhoff (time) average as

$$f^*(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f \circ \phi^t(x) dt$$

The Birkhoff theorem states that:

Proposition 7. Given a measure preserving dynamical system (M, μ, T) with $\mu(M) = 1$, if $f \in L^1(M, \mu)$, then

- the limit f^* exists almost everywhere
-

$$\int_M f^* d\mu = \int_M f d\mu$$

Definition 11. A measure preserving dynamical system is ergodic if and only if for every $f \in L^1$ exists c such that $f^*(x) = c$ for almost all values of x .

Proposition 8. *The following statements are equivalent:*

- $\forall f \in L^1, f^* = \text{const}$
- $\forall f \in L^1, f^* = \int f d\mu$
- if $f \in L^1$ is invariant, then it is constant
- if $A \subseteq M$, which we take to be measurable, is invariant, the measure is $\mu(A) \in \{0, 1\}$

Proof. • (1) implies (2) because, if f^* is constant, by Birkhoff's theorem it's integral is equal to the integral of f . Since f^* is constant, we can take it out of the integral, and

$$\int d\mu = 1$$

because μ is a probability measure. So

$$f^* = \int f d\mu$$

- (2) implies (3) because, if we take f invariant, so $f \circ T = f$, we can also say that $f \circ T^k = f$ for every k , which means that it is invariant at all times. This means that in calculating f^* we are summing f a number of times and then dividing by that number, so it remains f . But if $f^* = f$, by (2) we have that

$$\int f d\mu = f$$

which assures that f is a constant.

- (3) implies (4) because, if we apply (3) to $f = 1_A$,
- (4) implies (1) because, if we claim that a function $f : M \rightarrow \mathbb{R}$ is constant almost everywhere only if exists a constant l such that $\mu(\{x \in M | f(x) \leq l\}) \in (0, 1)$, this set is increasing as we increase l . Now, if l goes to $-\infty$ the measure goes to 0, and if l goes to ∞ , the measure goes to 1. The function can only jump in going from 0 to 1, but it can't do it all in one jump, so it must be constant except on the jumps.

□