MATH.APP.790 : Topics in Mathematics, Nonlinear time series analysis

Narayan P. Subramaniyam

Tampere University (MET Faculty)

January 2021

Summary so far

- Why nonlinear time series analysis ?
- Concept of phase space, attractor and chaos.
- Reconstruction of phase space from time series data.
- Estimation of embedding dimension and lag.
- The concept of dynamic invariant (e.g., the Lyapunov exponent)

Lecture overview

- The idea of (fractal) dimension.
- Information theoretic measures.
- Poincare's recurrence theorem .
- References

1. Kantz, Holger, and Thomas Schreiber. Nonlinear time series analysis. Vol. 7. Cambridge university press, 2004.
2. Puthanmadam Subramaniyam, Narayan. "Recurrence network analysis of EEG signals: A Geometric Approach." (2016).
3. Young, Lai-Sang. "Dimension, entropy and Lyapunov exponents." Ergodic theory and dynamical systems 2.1 (1982): 109-124.
4. J.L. Kaplan and J.A. Yorke, Lecture Notes in Math 730, 204 (1979)

The idea of dimension

What is dimension ? How to measure it ?

The idea of dimension

What is dimension ? How to measure it ?

Figure: What happens if you continually bisect a line ?

The idea of dimension

What is dimension ? How to measure it ?

Figure: What happens if you continually bisect a line ?

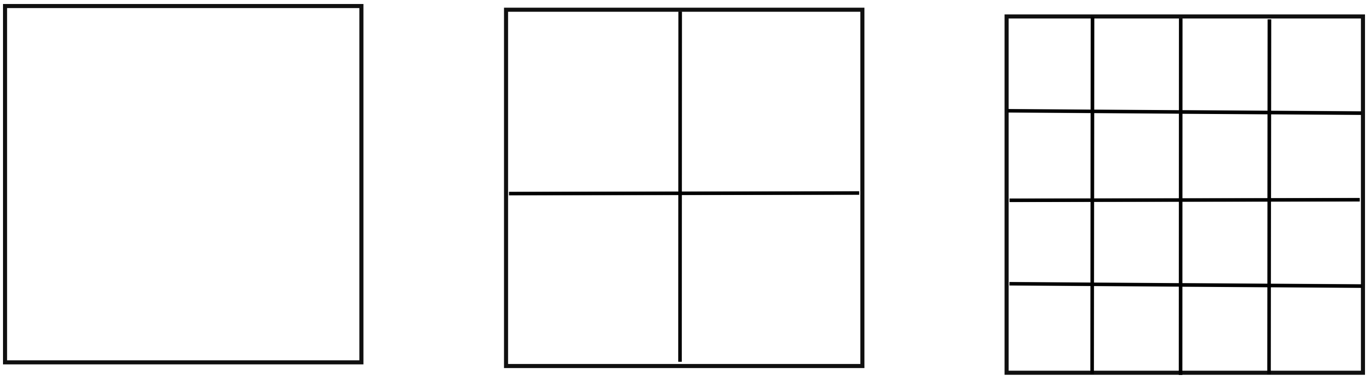


Figure: What happens if you continually bisect a square ?

The idea of dimension

What is dimension ? How to measure it ?

The idea of dimension

What is dimension ? How to measure it ?

Figure: Next iteration is made up of two $1/2$-sized copies of previous level.

The idea of dimension

What is dimension ? How to measure it ?

Figure: Next iteration is made up of two $1/2$-sized copies of previous level.

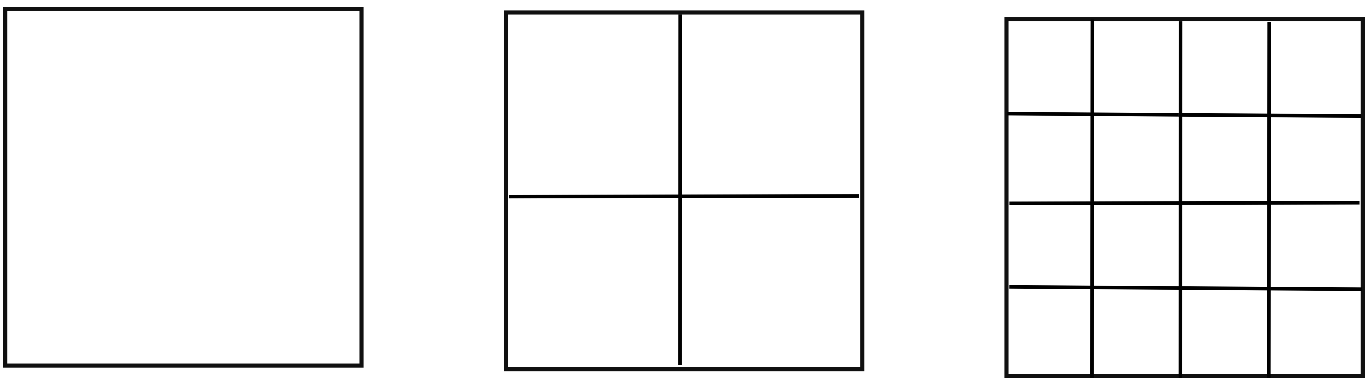


Figure: Next iteration is made up of four $1/4$-sized copies of previous level.

The idea of dimension

- Instead of bisecting, what happens if we continually trisect line (one-dimension) and a square (two-dimension) ?
- In case of the line : Each of the next iteration will have three $1/3$-sized copies of previous level.
- In case of the square : Each of the next iteration will have nine $1/9$-sized copies of previous level.
- How about a continually trisecting a cube ?

The idea of dimension

- In general, continually M-secting a object makes M^D copies of the previous level
- Here D is the dimension of the object.

The idea of dimension

- In general, continually M-secting a object makes M^D copies of the previous level
- Here D is the dimension of the object.

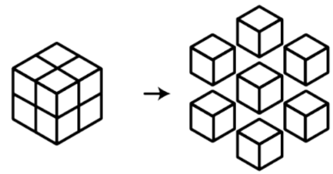
Given a D-dimensional object, create a geometric structure by repeatedly dividing the length of its sides by a number M.

The idea of dimension

- In general, continually M-secting a object makes M^D copies of the previous level
- Here D is the dimension of the object.

Given a D-dimensional object, create a geometric structure by repeatedly dividing the length of its sides by a number M.

- Each level is then made up of M^D copies of previous level
- Let N be the number of copies.
- $N = M^D$ or $\log N = D \log M$.
- Thus, $D = \log N / \log M$, i.e., dimension is defined as ratio between (logarithm) of number of copies we get (N) and (logarithm) amount by which we reduce the length of the sides (M).



(a) Graphical representation

Object	Scale s	Pieces a
Line	$1/2$ $1/n$	2 n
Square	$1/2$ $1/n^2$	4 n^2
Cube	$1/2$ $1/n^2$	8 n^3

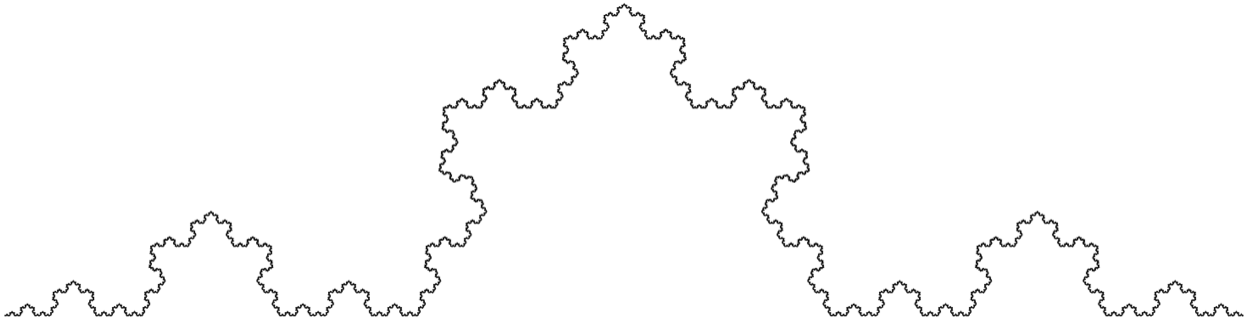
(b) Table of the scaling factors

The idea of fractal dimension

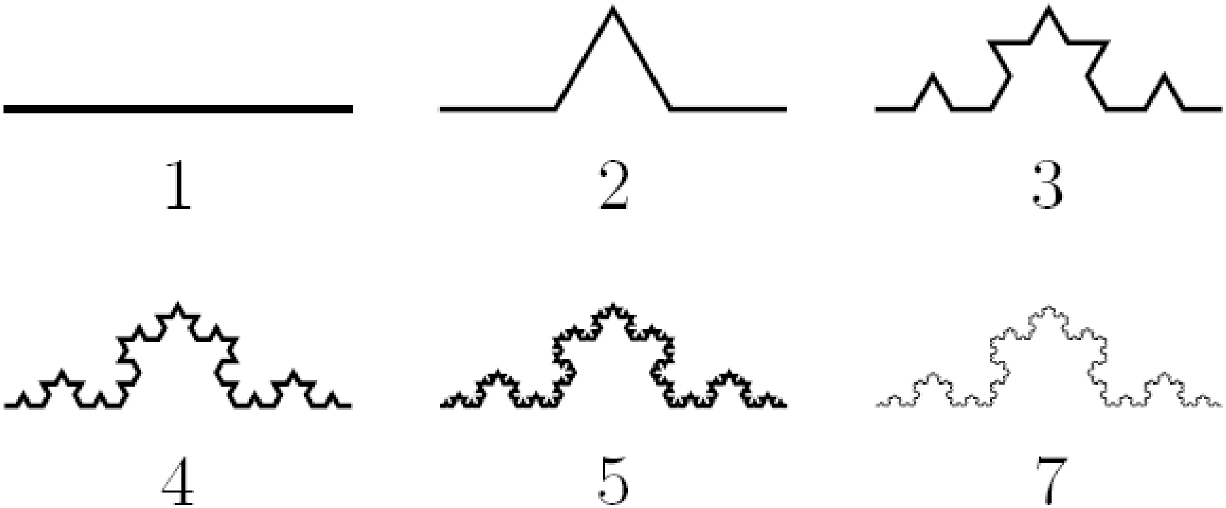
- Fractal geometry deals with non-idealized, crinkly objects that are characterized by their fractional dimensions.
- Attractors can be considered as conceptual objects, that arise in the phasee space of dissipative chaotic (and in some cases even non-chaotic) systems.
- Attractors with seemingly complicated geometry are known as strange attractors, which are characterized by fractional dimension (e.g., Lorenz attractor).
- Fractional dimension is usually exhibited by objects that display unusual kind of self-similarity to some degree.
- This self-similarity is sometimes exact, but more often it is just approximate or statistical!

Dimension of the Koch curve

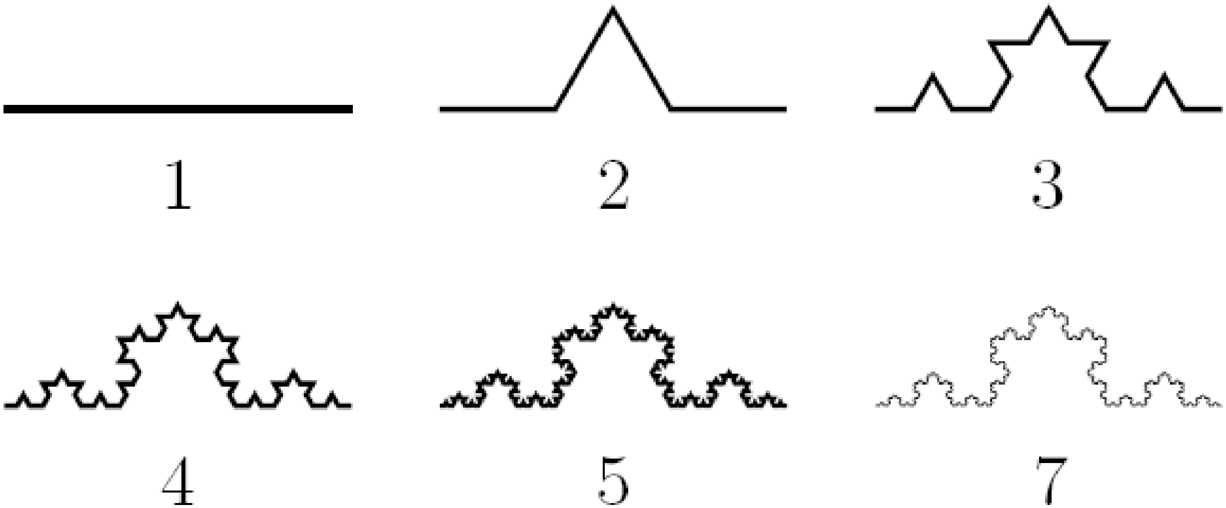
What is the dimension of this object ? 1 or 2 or something in between ....



Dimension of the Koch curve

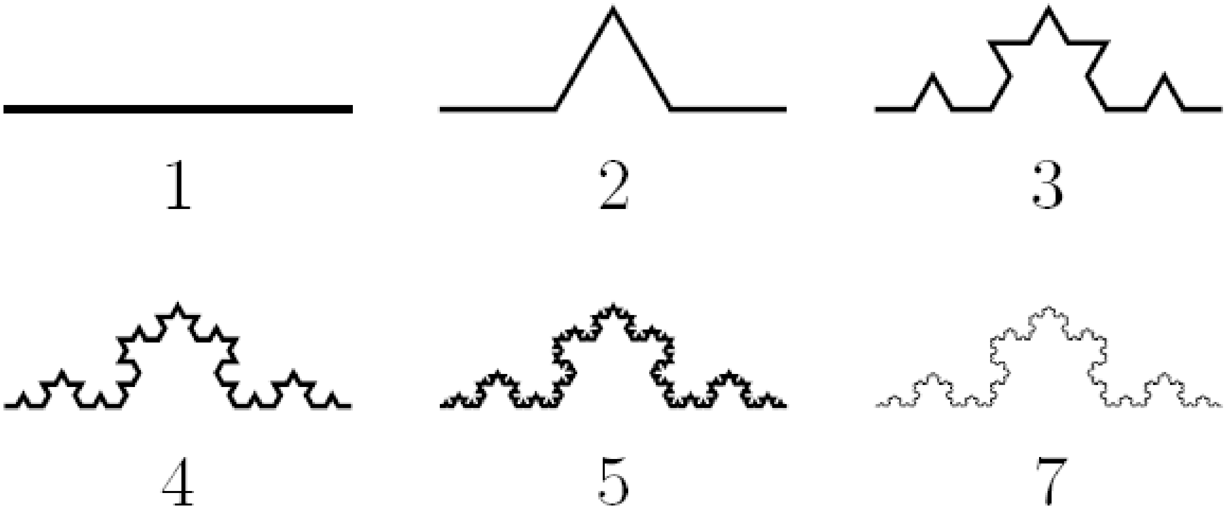


Dimension of the Koch curve



Here $M = 3$, i.e. we are reducing the length of each side by 3 and $N = 4$, i.e. the number of copies in the next level is 4 times the previous level.

Dimension of the Koch curve



Here $M = 3$, i.e. we are reducing the length of each side by 3 and $N = 4$, i.e. the number of copies in the next level is 4 times the previous level.

$$D = \log 3 / \log 4 \approx 1.26$$

Fractal dimensions

- This (non-integer) number tells us how do the number of copies scale with the decrease in the size of the segment.
- Density of the self-similarity.
- Examples of fractals

Dimension of an attractor : Box-counting dimension

- The dimension of an attractor is a dynamic invariant.

Dimension of an attractor : Box-counting dimension

- The dimension of an attractor is a dynamic invariant.
- Consider a k-dimensional phase space, comprising of N non-empty hypercubes of length l, that is needed to fully cover the attractor embedded in the phase space.

Dimension of an attractor : Box-counting dimension

- The dimension of an attractor is a dynamic invariant.
- Consider a k-dimensional phase space, comprising of N non-empty hypercubes of length l, that is needed to fully cover the attractor embedded in the phase space.
- The number of hypercubes needed to cover the attractor typically scales as a function of the length parameter.
- $N(l) \approx l^{-D_0}$, where D_0 can be thought of as the dimension of attractor the hypercubes are covering.

Dimension of an attractor : Box-counting dimension

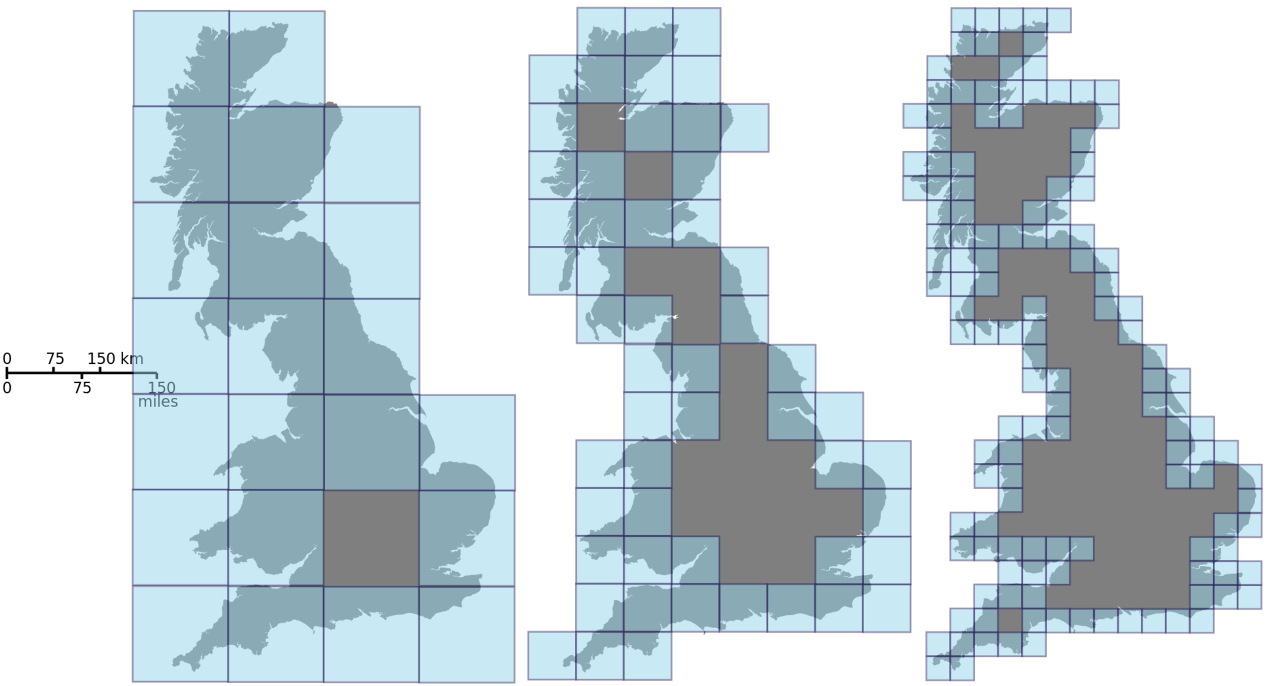
- The dimension of an attractor is a dynamic invariant.
- Consider a k-dimensional phase space, comprising of N non-empty hypercubes of length l, that is needed to fully cover the attractor embedded in the phase space.
- The number of hypercubes needed to cover the attractor typically scales as a function of the length parameter.
- $N(l) \approx l^{-D_0}$, where D_0 can be thought of as the dimension of attractor the hypercubes are covering.
- Thus the dimension D_0 of the attractor can be approximated as follows,

$$D_0 = \lim_{l \to 0} \frac{\log N(l)}{\log l}.$$

Dimension of an attractor : Box-counting dimension

- Theoretically, the largest value of D_0 as $l \to 0$, is k (phase space dimension).
- Also, $D_0 \approx k$ if the embedding dimension is inadequate to unfold the attractor (as the attractor fills up the phase space due to projection).
- Computing D_0 for increasing embedding dimensions until a saturation in the value of D_0 is observed might indicate that the attractor has been properly unfolded.

Coast of Great Britan



More notions on dimension

Suppose again that $N(l)$ is the number of hypercubes needed to cover an attractor set. If the number of points of the attractor set contained in the i^{th} hypercube is N_i, then the probability p_i for a randomly chosen point of the attractor set to be in the i^{th} hypercube is $p_i = N_i/N(l)$. Now one can define the dimension spectrum as

$$D_q = \lim_{l \to 0} \frac{1}{q-1} \frac{\log N(l)}{\log l}.$$

More notions on dimension

Suppose again that $N(l)$ is the number of hypercubes needed to cover an attractor set. If the number of points of the attractor set contained in the i^{th} hypercube is N_i, then the probability p_i for a randomly chosen point of the attractor set to be in the i^{th} hypercube is $p_i = N_i/N(l)$. Now one can define the dimension spectrum as

$$D_q = \lim_{l \to 0} \frac{1}{q-1} \frac{\log N(l)}{\log l}.$$

1. For $q = 0$, we get the box-counting dimension.

More notions on dimension

Suppose again that $N(l)$ is the number of hypercubes needed to cover an attractor set. If the number of points of the attractor set contained in the i^{th} hypercube is N_i, then the probability p_i for a randomly chosen point of the attractor set to be in the i^{th} hypercube is $p_i = N_i/N(l)$. Now one can define the dimension spectrum as

$$D_q = \lim_{l \to 0} \frac{1}{q-1} \frac{\log N(l)}{\log l}.$$

1. For $q = 0$, we get the box-counting dimension.
2. For $q = 1$, we get the information dimension.

More notions on dimension

Suppose again that $N(l)$ is the number of hypercubes needed to cover an attractor set. If the number of points of the attractor set contained in the i^{th} hypercube is N_i, then the probability p_i for a randomly chosen point of the attractor set to be in the i^{th} hypercube is $p_i = N_i/N(l)$. Now one can define the dimension spectrum as

$$D_q = \lim_{l \to 0} \frac{1}{q-1} \frac{\log N(l)}{\log l}.$$

1. For $q = 0$, we get the box-counting dimension.
2. For $q = 1$, we get the information dimension.
3. For $q = 2$, we get the correlation dimension.

Dimension of an attractor : Information dimension

$$D_1 = \lim_{l \to 0} \frac{1}{q-1} \frac{\log \sum_i p_i p_i^{q-1}}{\log l},$$

Dimension of an attractor : Information dimension

$$D_1 = \lim_{l \to 0} \frac{1}{q-1} \frac{\log \sum_i p_i p_i^{q-1}}{\log l} \approx \lim_{l \to 0} \frac{1}{q-1} \frac{\log \sum_i p_i [1 + (q-1) \log p_i]}{\log l}$$

Dimension of an attractor : Information dimension

$$D_1 = \lim_{l \to 0} \frac{1}{q-1} \frac{\log \sum_i p_i p_i^{q-1}}{\log l} \approx \lim_{l \to 0} \frac{1}{q-1} \frac{\log \sum_i p_i [1 + (q-1) \log p_i]}{\log l}$$

so that,

$$D_1 = \lim_{l \to 0} \frac{\sum_{i=1}^{N(l)} p_i \log p_i}{\log l},$$

Dimension of an attractor : Information dimension

$$D_1 = \lim_{l \to 0} \frac{\sum_{i=1} p_i \log p_i}{\log l},$$

- Information dimension tells us how information scales with box size.
- To calculate D_1, each box is weighted with the measure.
- Thus, information dimension characterizes most directly the measure of the attractor.

Dimension of an attractor : Correlation dimension

For $q = 2$, we have

$$D_2 = \lim_{l \to 0} \frac{\log \sum_i p_i^2}{\log l}$$

Dimension of an attractor : Correlation dimension

For $q = 2$, we have

$$D_2 = \lim_{l \to 0} \frac{\log \sum_i p_i^2}{\log l}$$

What is $\sum_i p_i^2$?

Dimension of an attractor : Correlation dimension

For $q = 2$, we have

$$D_2 = \lim_{l \to 0} \frac{\log \sum_i p_i^2}{\log l}$$

- The probability that two points lie within cells of the length l is $\sum_i p_i^2$.
- This scales in the same way as the probability that two points in the data set are separated by a distance less than l.
- This can be determined by correlation function [Grassberger and Procaccia algorithm].

$$C(l) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j}^{N} \Theta(l - |x_i - x_j|) \tag{1}$$

Dimension of an attractor : Correlation dimension

Correlation function

$$C(l) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j}^{N} \Theta(l - |x_i - x_j|) \tag{2}$$

- Where x_i and x_j are two points on the attractor and $|x_i - x_j|$ is the distance (for example Euclidean) between them.
- $\Theta(\cdot)$ is known as the Heaviside function which takes the value of 1 when this distance is less than the threshold l, and 0 otherwise.

For sufficiently large N and small l, we have, $C(l) \approx r^{D_2}$ and thus the correlation dimension can then be given as,

$$D_2 = \lim_{l\to 0} \frac{\log C(l)}{\log l}$$

Pitfalls in computing D_2

- D_2 can be biased due to autocorrelation in the time series.
- This can be reduced to some extent by discarding pair of points in the trajectories with time indices less than the autocorrelation time. (e.g., choose $\tau >$ 3 times the autocorrelation time).
- It has been shown that insufficient data length can also bias the estimate of D_2 (The bound for D_2 is $2\log_{10} N$).
- Presence of noise can lead to overestimation of D_2

Information theory and dynamical systems

What is Entropy ?

Information theory and dynamical systems

What is Entropy ?

Consider a sequence of N coin tosses. There are 2^N possible outcomes. The concept of entropy can be used to measure the uncertainty in the outcome,

$$S = \log 2^N$$

Information theory and dynamical systems

- The information capacity is also given by $I = S = \log 2^N$.
- Measurement of a particular sequence of H and T gives information about the system ($N \log 2$ bits of information).
- Generalizing to a system of N possible results with independent probabilities p_i, we can define entropy as,

$$I = -\sum_i p_i \log p_i$$

Entropy and dynamical systems

- Consider an attractor set $\mathcal{A}$, partitioned into N boxes, $\mathcal{B} = \{B_i\}$, $i = 1 \dots N$.
- $B_1, B_2, \cdots, B_N \subset \mathcal{A}$, where $B_i \neq \emptyset$ and $B_i \cap B_j = \emptyset$ for all i,j
- The probability of finding a point in the box B_i is $p_i = \int_{B_i} \rho(x) dV_i$.
- The entropy of the dynamical system is defined as,

$$S = -\sum_i p_i \log p_i$$

  where $\rho(x)$ is the invariant measure and p_i is given as the integral over this measure over the ith element.

With the above definition, we can compute the information dimension as discussed before.

Entropy and dynamical systems

- Consider an attractor set $\mathcal{A}$, partitioned into N boxes, $\mathcal{B} = \{B_i\}$, $i = 1 \dots N$.
- $B_1, B_2, \cdots, B_N \subset \mathcal{A}$, where $B_i \neq \emptyset$ and $B_i \cap B_j = \emptyset$ for all i,j
- The probability of finding a point in the box B_i is $p_i = \int_{B_i} \rho(x) dV_i$.
- The entropy of the dynamical system is defined as,

$$S = -\sum_i p_i \log p_i$$

  where $\rho(x)$ is the invariant measure and p_i is given as the integral over this measure over the ith element.

With the above definition, we can compute the information dimension as discussed before.

Another quantity, known as Kolmogorov-Sinai entropy can also be defined, which quantifies how uncertainty or information of the system evolves in time.

Kolmogorov-Sinai entropy

Let $P(B_{i_0}, B_{i_1}, \cdots, B_{i_T})$ be the probability that a trajectory is in the set B_{i_j} and time $n = j$. Let

$$K_T = -\sum_{i_0, i_1, \cdots, i_T} P(B_{i_0}, B_{i_1}, \cdots, B_{i_T}) \log_2 P(B_{i_0}, B_{i_1}, \cdots, B_{i_T})$$

Now, consider the difference $K_{T+M} - K_T$. Intuitively, this is the information to correctly identify the hypercubes $B_{T+1}, B_{T+2}, \cdots, B_{T+M}$ visited by a trajectory between times T and $T + M$. Now, one can define the Kolmogorov entropy as

$$K \equiv \lim_{max_i size(B_i) \to 0} [\lim_{T \to \infty} (K_{T+1} - K_T)]$$

For a chaotic system $K_T > 0$ and periodic system $K_T = 0$.

Kolmogorov-Sinai entropy

Let $P(B_{i_0}, B_{i_1}, \cdots, B_{i_T})$ be the probability that a trajectory is in the set B_{i_j} and time $n = j$. Let

$$K_T = -\sum_{i_0, i_1, \cdots, i_T} P(B_{i_0}, B_{i_1}, \cdots, B_{i_T}) \log_2 P(B_{i_0}, B_{i_1}, \cdots, B_{i_T})$$

Now, consider the difference $K_{T+M} - K_T$. Intuitively, this is the information to correctly identify the hypercubes $B_{T+1}, B_{T+2}, \cdots, B_{T+M}$ visited by a trajectory between times T and $T + M$. Now, one can define the Kolmogorov entropy as

$$K \equiv \lim_{max_i size(B_i) \to 0} [\lim_{T \to \infty} (K_{T+1} - K_T)]$$

For a chaotic system $K_T > 0$ and periodic system $K_T = 0$. Also, it can be shown that Kolmogorov-Sinai entropy is bounded by the sum of positive exponents : $K <= \sum_{\lambda_i > 0} \lambda_i$

Poincaré recurrences

- The metatheorem of dynamical systems theory states that, for an appropriately bounded phase space X, the trajectories of the motion will exhibit some form of recurrence.
- That is, they will return close to their initial position.
- In 1890, Poincaré formulated the first precise result and proved that whenever a dynamical system preserves volume, almost all the trajectories return arbitrarily close to their initial position, infinite number of times

Poincaré recurrence theorem

Formally, let $(X, \mathcal{B}, \mu)$ be a measure space of finite total measure. Here, $\mathcal{B}$ is a σ-algebra (i.e., $\mathcal{B} \subset 2^X$) μ is a probability measure. Let $f : X \to X$ be a function such that $f^{-1}\mathcal{A} \in \mathcal{B}$ for any $\mathcal{A} \in \mathcal{B}$ and $\mu(f^{-1}\mathcal{A}) = \mu(\mathcal{A})$. Such a measure μ is called f-invariant.

Theorem

Let $f : (X, \mathcal{B}) \to (X, \mathcal{B})$ and $(X, \mathcal{B}, \mu)$ be a f-invariant measure with $\mu(X) < \infty$. For $\mathcal{A} \in \mathcal{B}$ with $\mu(\mathcal{A}) > 0$ almost every $x \in \mathcal{A}$ has the property that :

$$\mu(\{x \in \mathcal{A} : \text{ there exists infinite positive integers } n \in \mathbb{Z}^+ \text{ such that } f^n x \in \mathcal{A}\}) = \mu(\mathcal{A}).$$

The set $\{f^n x \,|\, n \in \mathbb{Z}^+\}$ is called the orbit of x and $x \in \mathcal{A}$ is **recurrent** if the orbit of x intersects $\mathcal{A}$ infinitely many times.

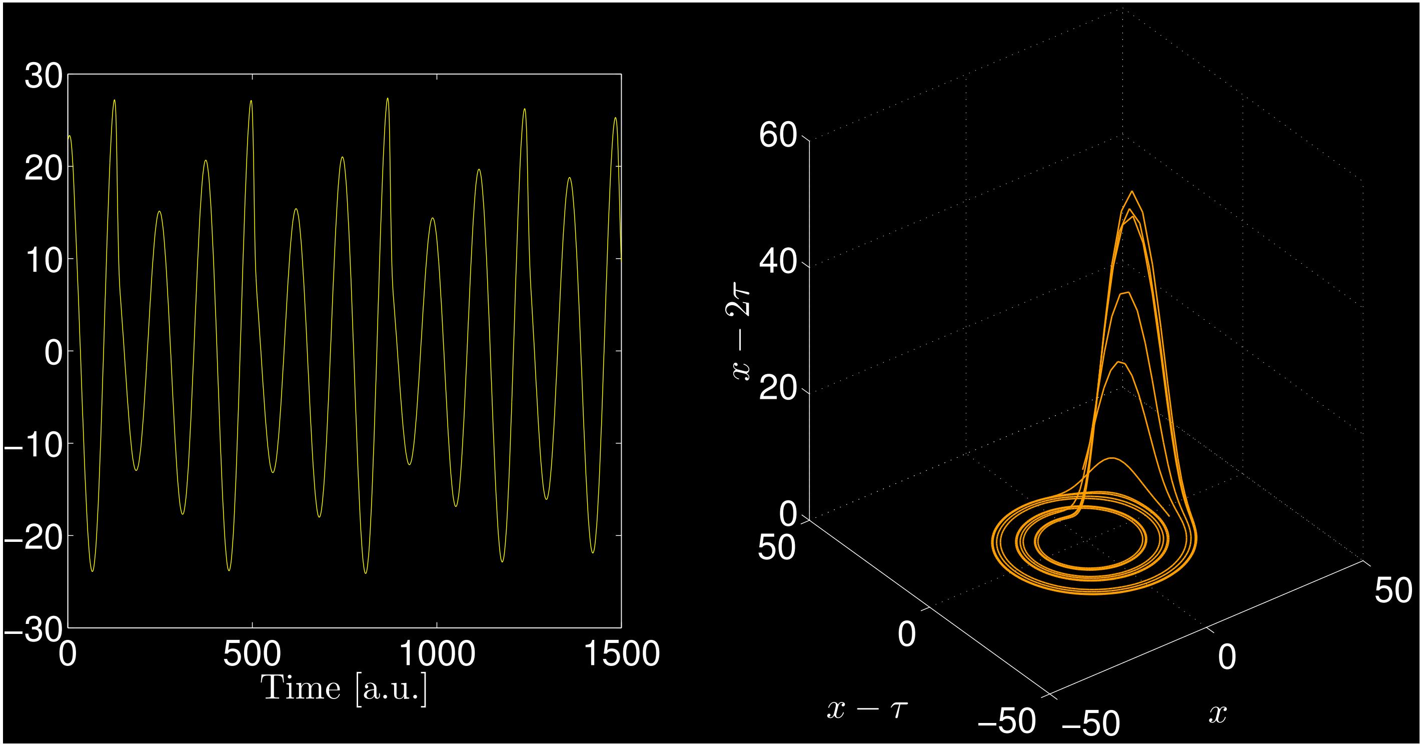
Poincaré recurrence theorem

Thus, the Poincaré recurrence theorem states that if μ is f-invariant then almost every point of a measurable subset with positive measure is recurrent.

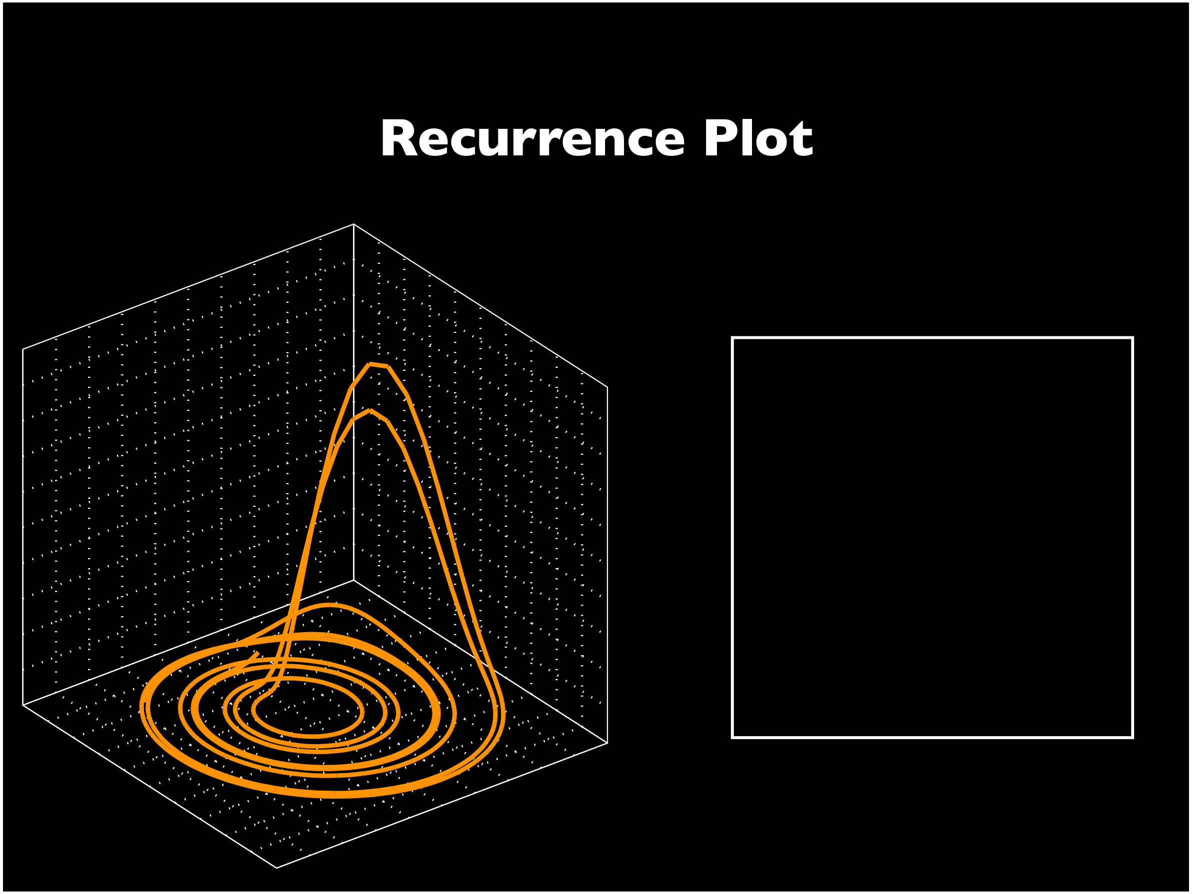
Recurrence plots and networks

- Recurrence is a fundamental property of a dynamical system.
- Recurrences in phase space can be easily visualized as **recurrence plots**
- The concept of recurrence can be reconsidered to define a complex network based on time series, known as a **recurrence network**

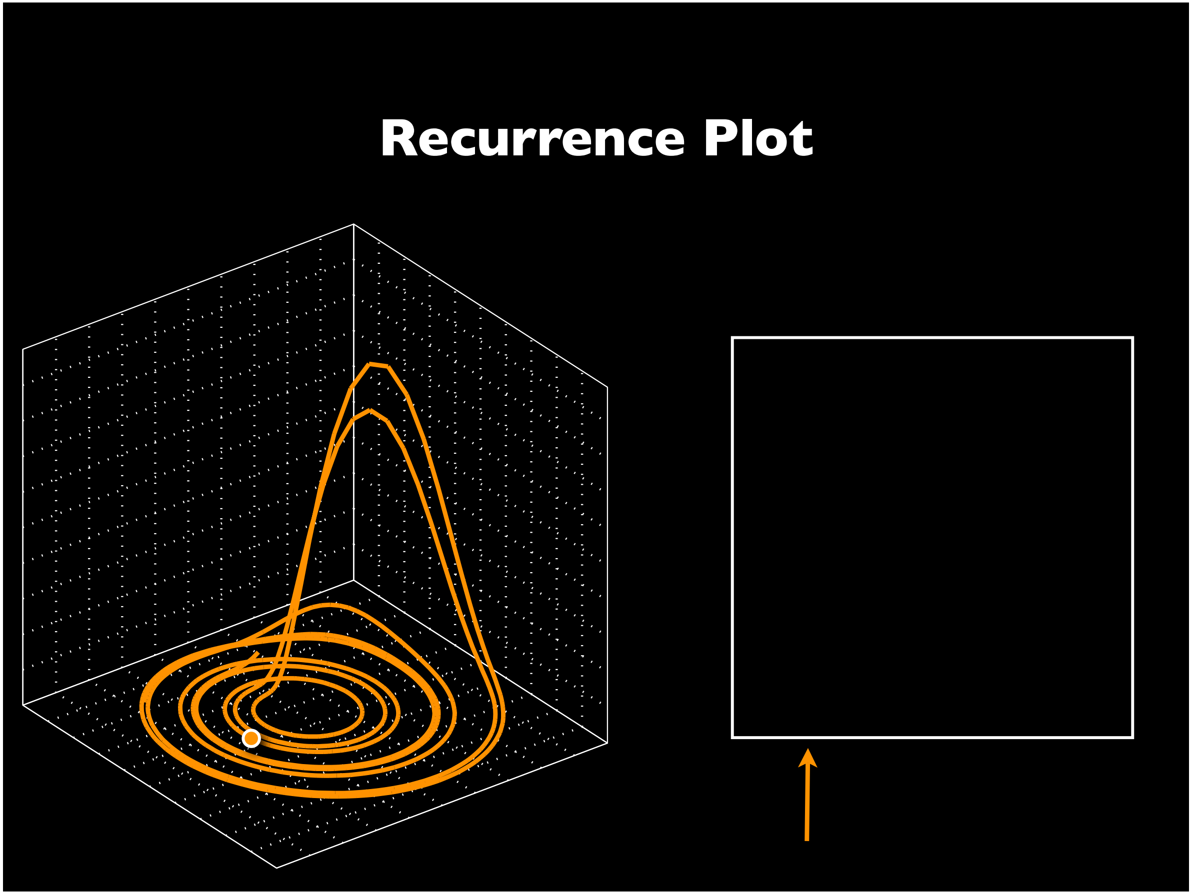
Time series to Phase Space



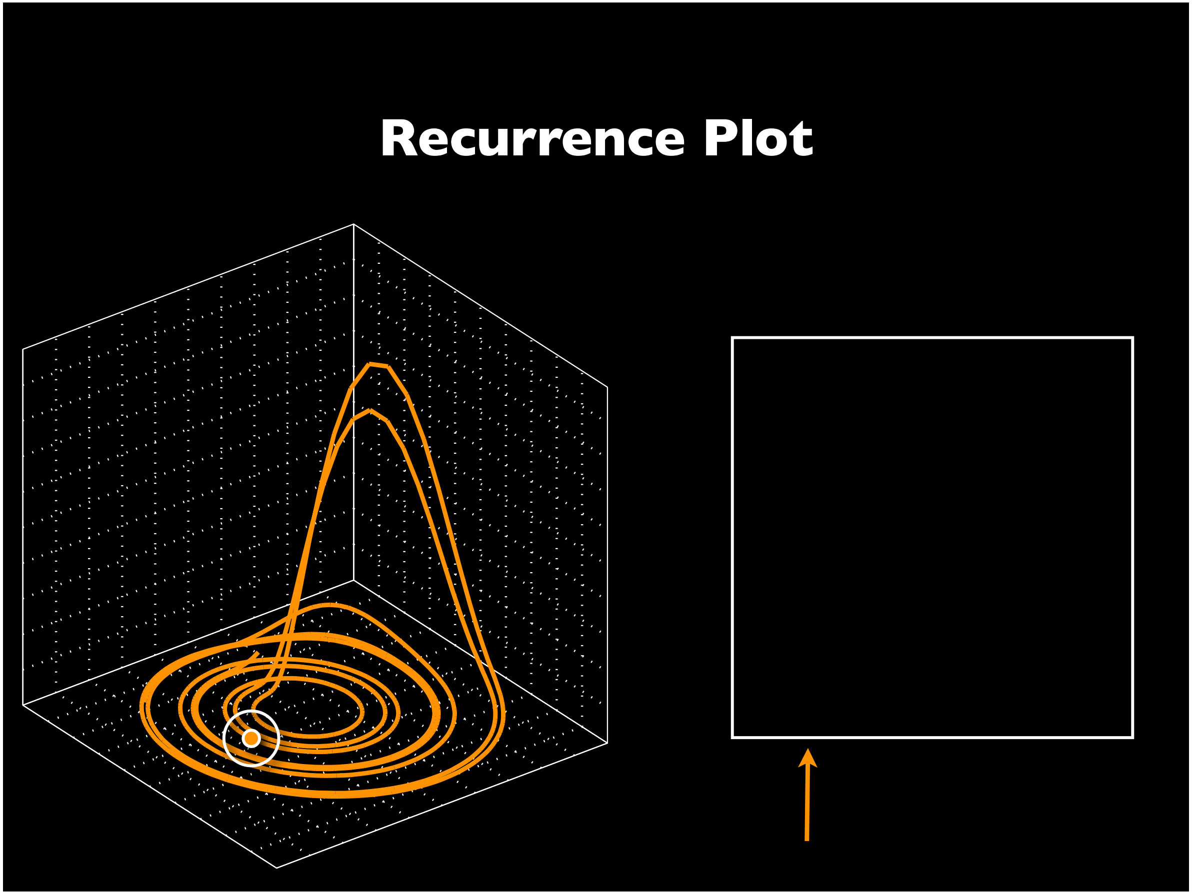
Recurrence plot



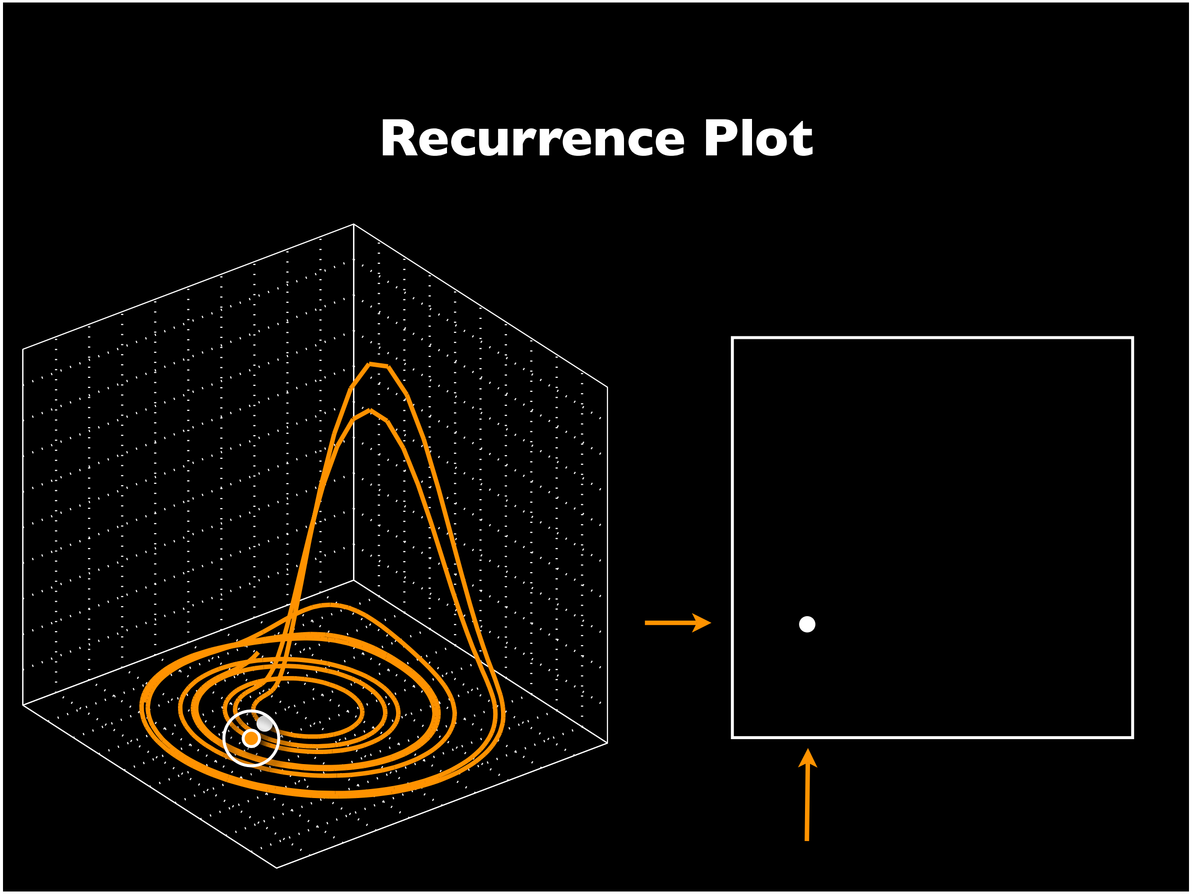
Recurrence plot



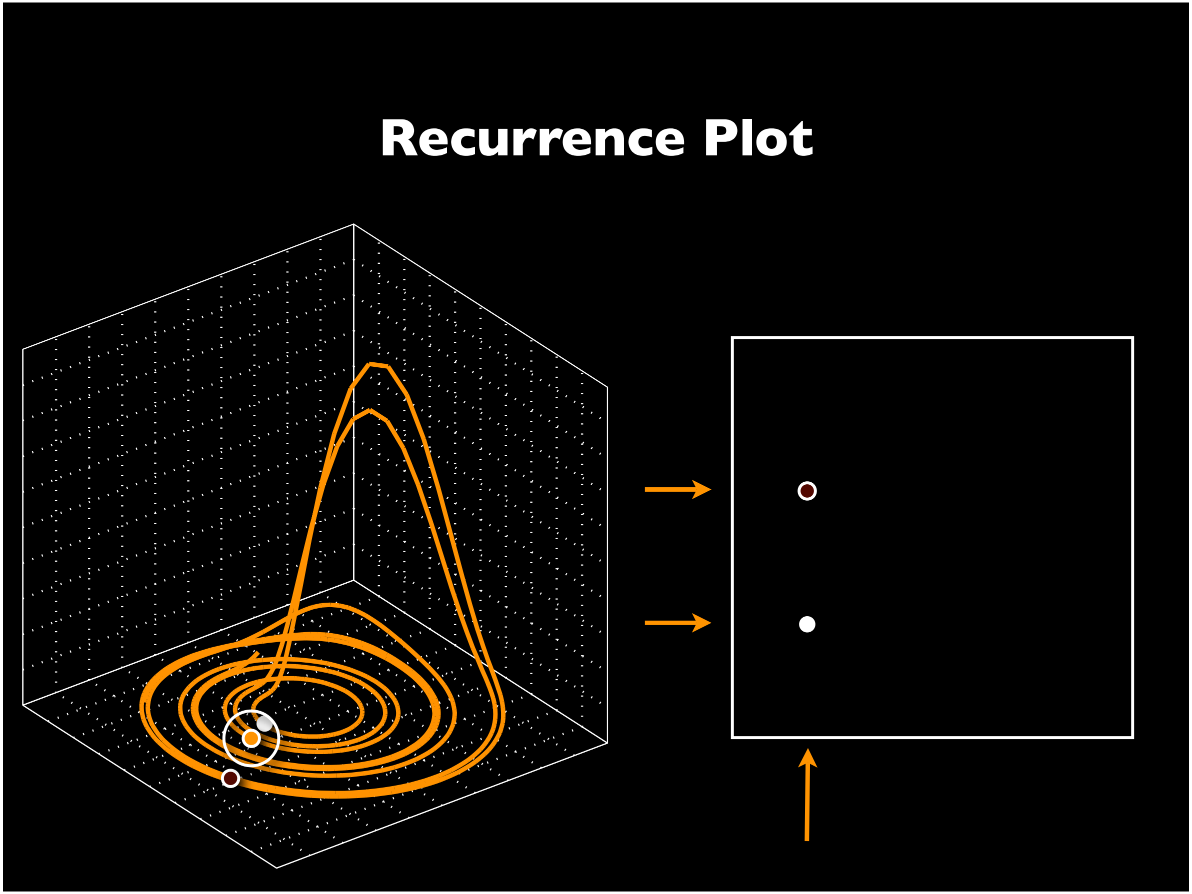
Recurrence plot



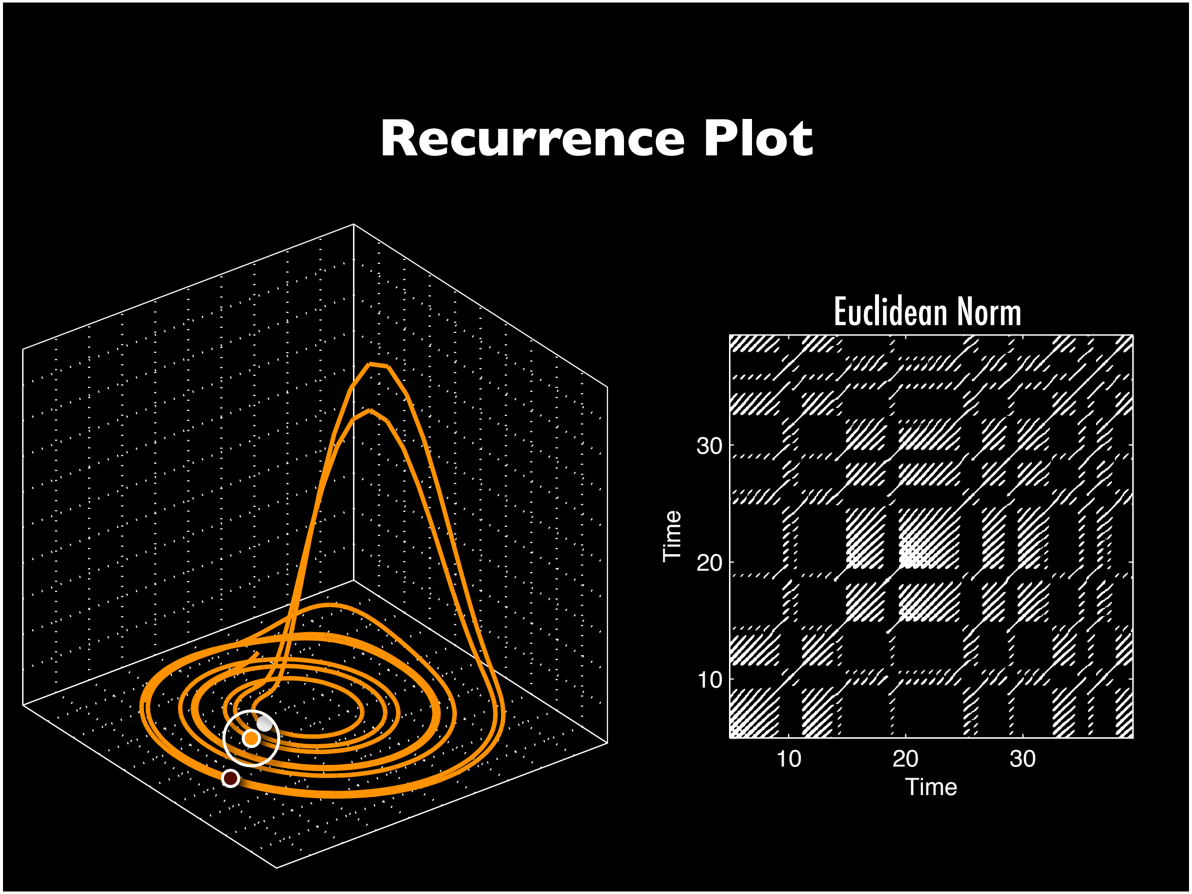
Recurrence plot



Recurrence plot



Recurrence plot



Recurrence Network