- We are going to deal in this chapter with chaotic dynamical systems, and more specifically with the estimation of some parameters characterizing these systems.
- As a matter of fact, if the search for real-life chaotic systems is a bit outdated, these parameters present a specific interest in many applications (physics, biomedical data analysis, finance, …).

- Estimation of these parameters aims at:
  - Detecting the presence of chaotic dynamics
  - Determining the dimension of the underlying mechanism
  - Quantifying the complexity of this dynamics
  - Obtaining features for classification purposes.

- There is no global definition of chaos. One sometimes speaks of the apparently stochastic evolution of a deterministic system, with an exponential sensitivity to initial conditions.

$$\frac{dX(t)}{dt} = G[X(t)] \qquad X(n) = F[X(n-1)]$$

- One speaks also of a bounded dynamics in equilibrium regime, which corresponds neither to a fixed point nor a limit cycle.

- One cannot have a chaotic dynamics with a linear system. The linear AR model:

$$x(n) = a_1 x(n-1) \cdots + a_p x(n-p) + \varepsilon(n)$$

- Can be cast in a Markov (state-space) representation:

$$\begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-p+1) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x(n-1) \\ x(n-2) \\ \vdots \\ x(n-p) \end{bmatrix} + \begin{bmatrix} \varepsilon(n) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

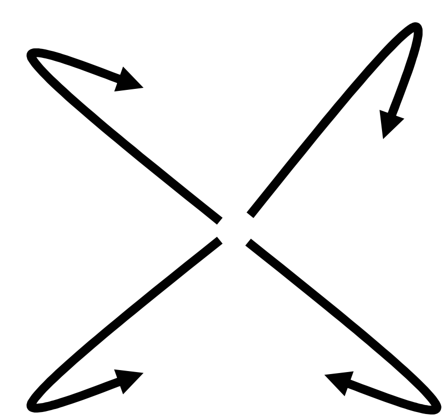
- To sum up:

$$X(n) = \mathbf{A}X(n\text{-}1) + E(n)$$

- If there is no excitation $E(n)$ three cases are possible:
  - Moduli of the eigenvalues of $\mathbf{A}$ are all < 1, ($\Leftrightarrow$ pole moduli < 1): $\|X(n)\|$ converges towards 0.
  - Some eigenvalues of $\mathbf{A}$ have a modulus >1, ($\Leftrightarrow$ pole moduli > 1): $\|X(n)\|$ grows without bound.
  - Neutrally stable case (atypical): limit cycle.

- In broad terms there are two cases:

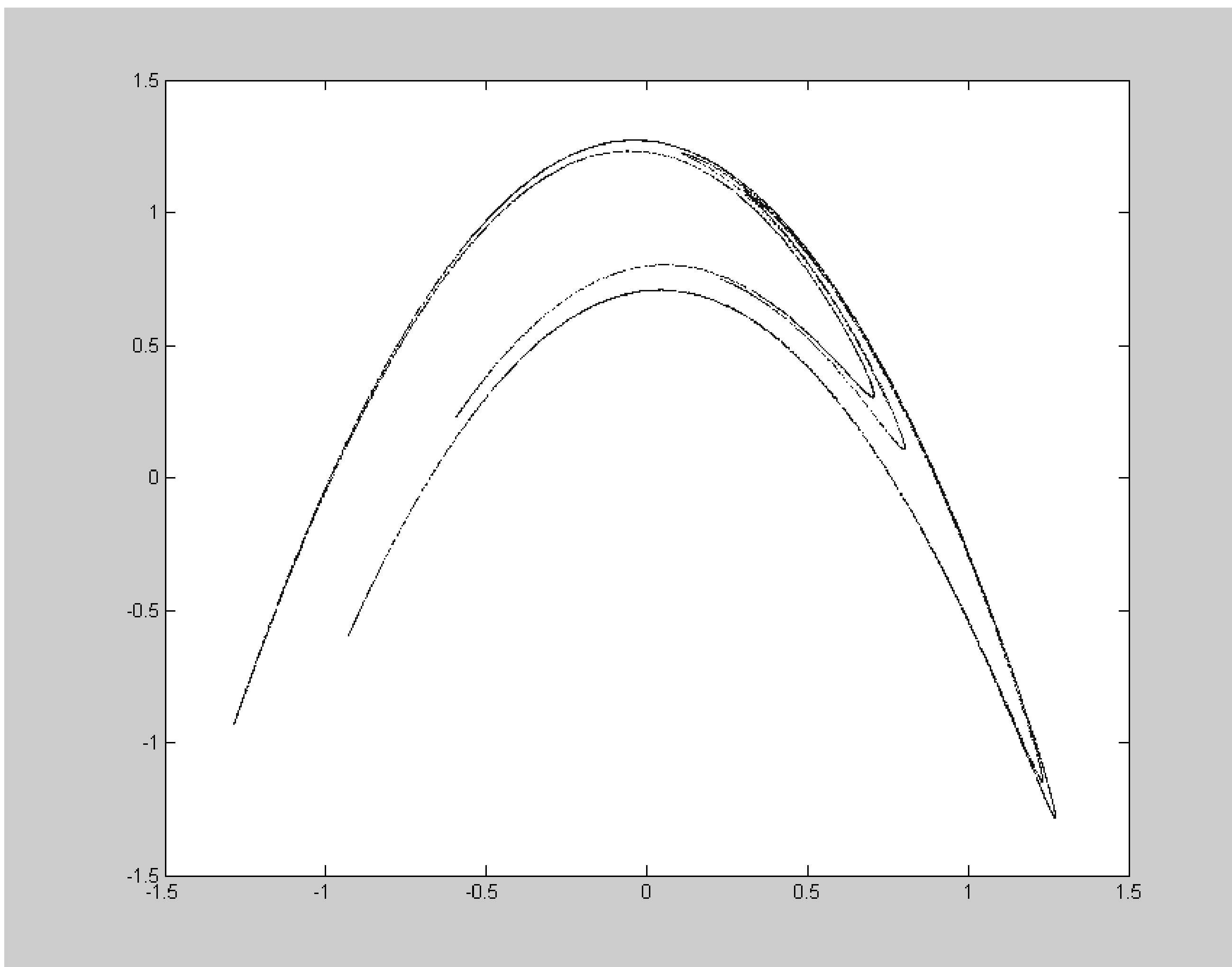
- But if the dynamics is nonlinear, it can “fold” the trajectories, so that it remains bounded:



- this succession of expansions/contractions coupled with the sensitivity to initial conditions is responsible for this aperiodic evolution.

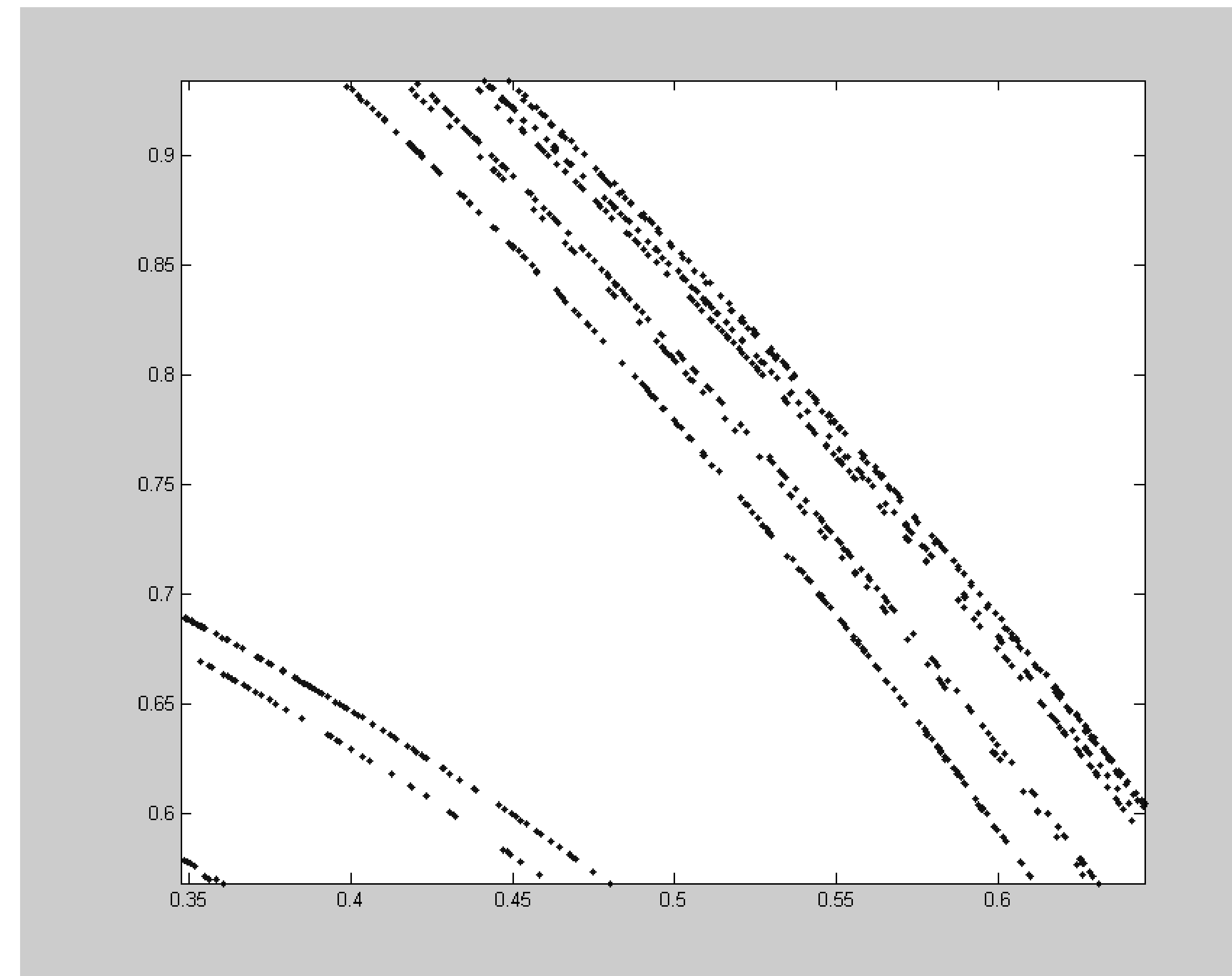
- Example: Hénon's Map:

$$\begin{bmatrix} x(n) \\ x(n-1) \end{bmatrix} = \begin{bmatrix} 1-1.4x(n-1)^2+0.3(x(n-2)) \\ x(n-1) \end{bmatrix} = F\left\{\begin{bmatrix} x(n-1) \\ x(n-2) \end{bmatrix}\right\}$$



$x(n)$

$x(n\text{-}1)$



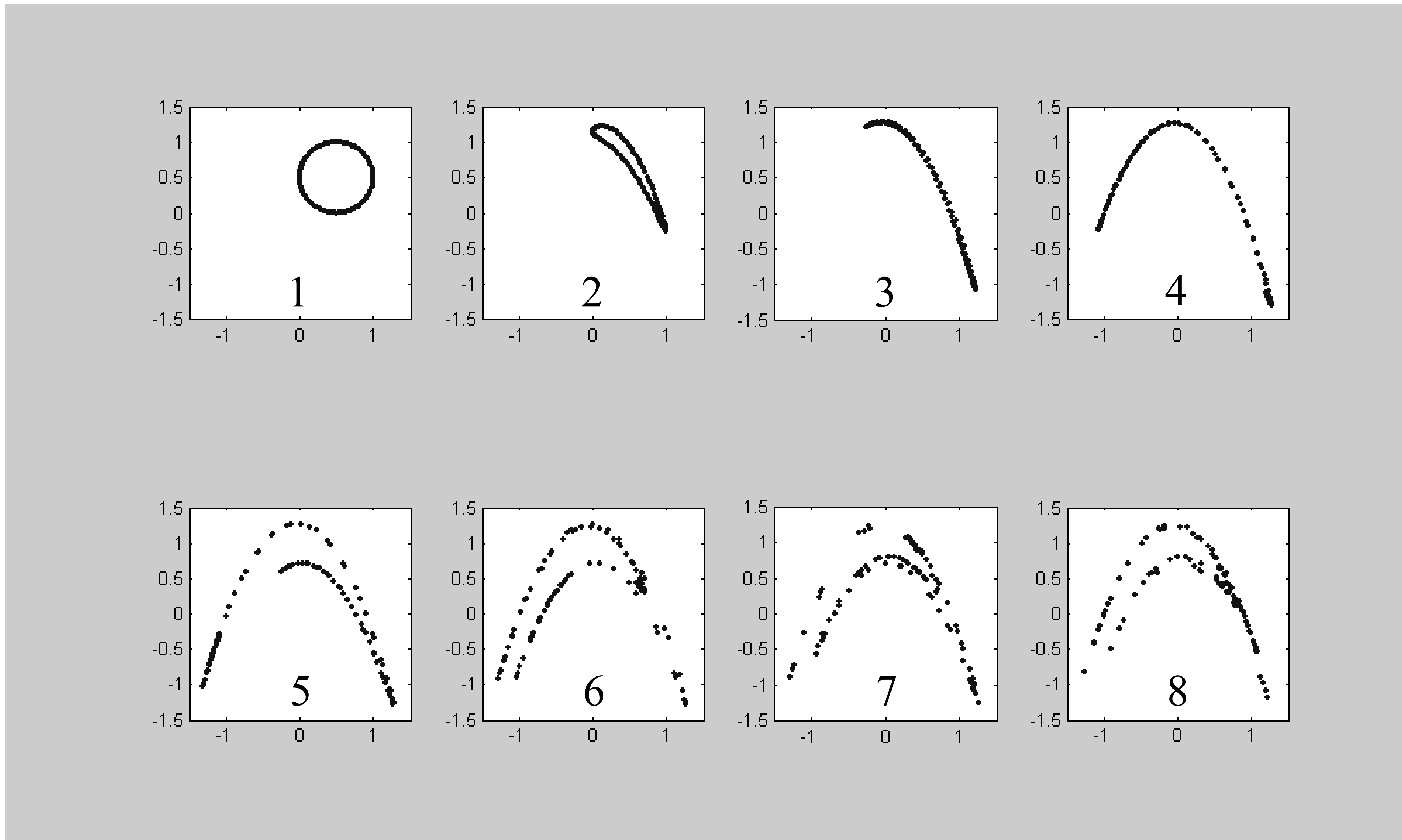
zoom

- Chaotic systems evolve generically towards a *strange attractor* characterized by:
  - A null volume
  - An exponentially fast separation of trajectories initially close
  - A dimension often fractal
  - An invariant measure ρ which enables the definition of mean values.

- This concept is linked to ergodicity: for an infinite number of initial conditions in the basin of attraction, the characteristics of the trajectories (such as point density in a region) are independent of the former.
- This is illustrated by applying Hénon's map simultaneously to a large number of points for several iterations. Successive images fill the attractor the same way a single trajectory would.

INVARIANT MEASURE (2)

- Example: iteration of 100 points on a circle



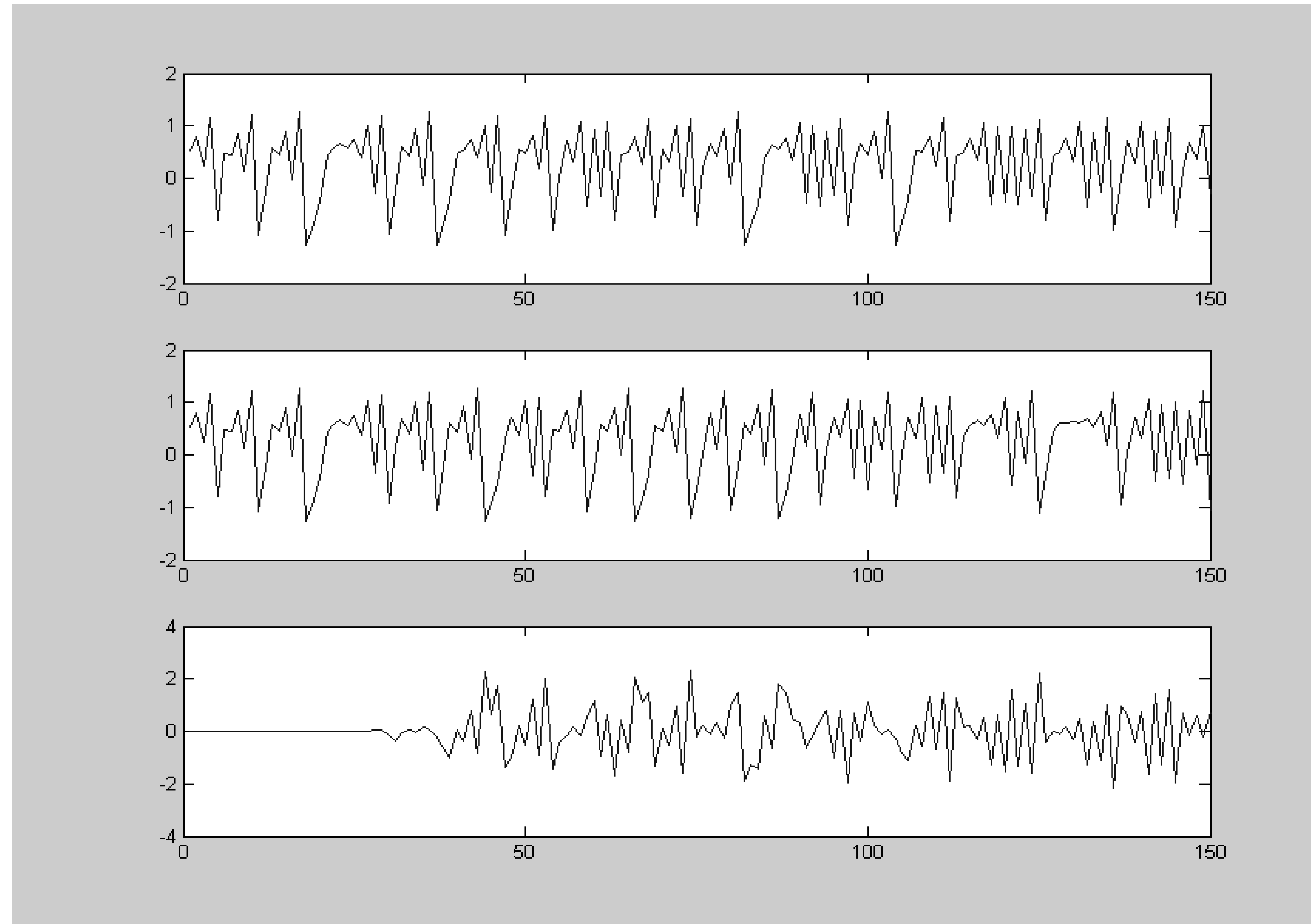
SEPARATION OF CLOSE TRAJECTORIES

- Sensitivity to initial conditions

$x(n)$

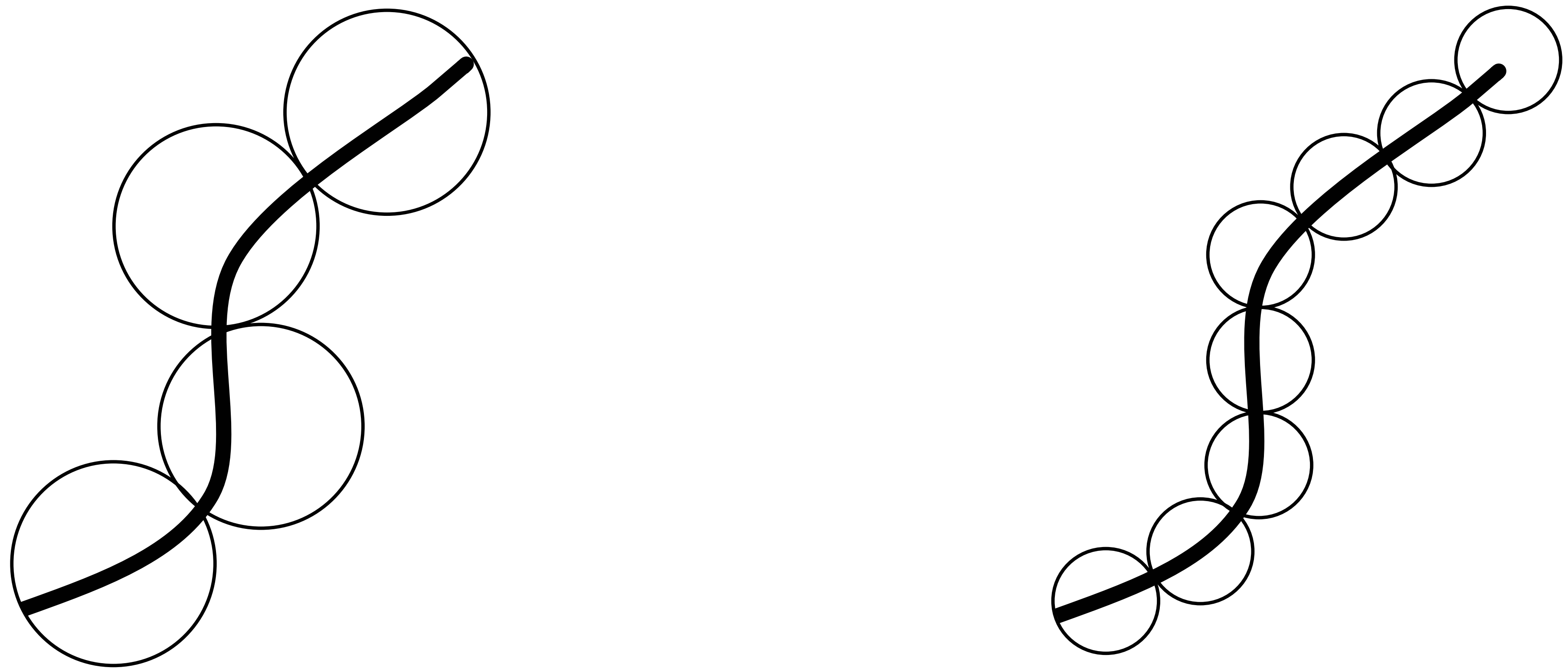
modification by 10^{-6} of one component

signal difference



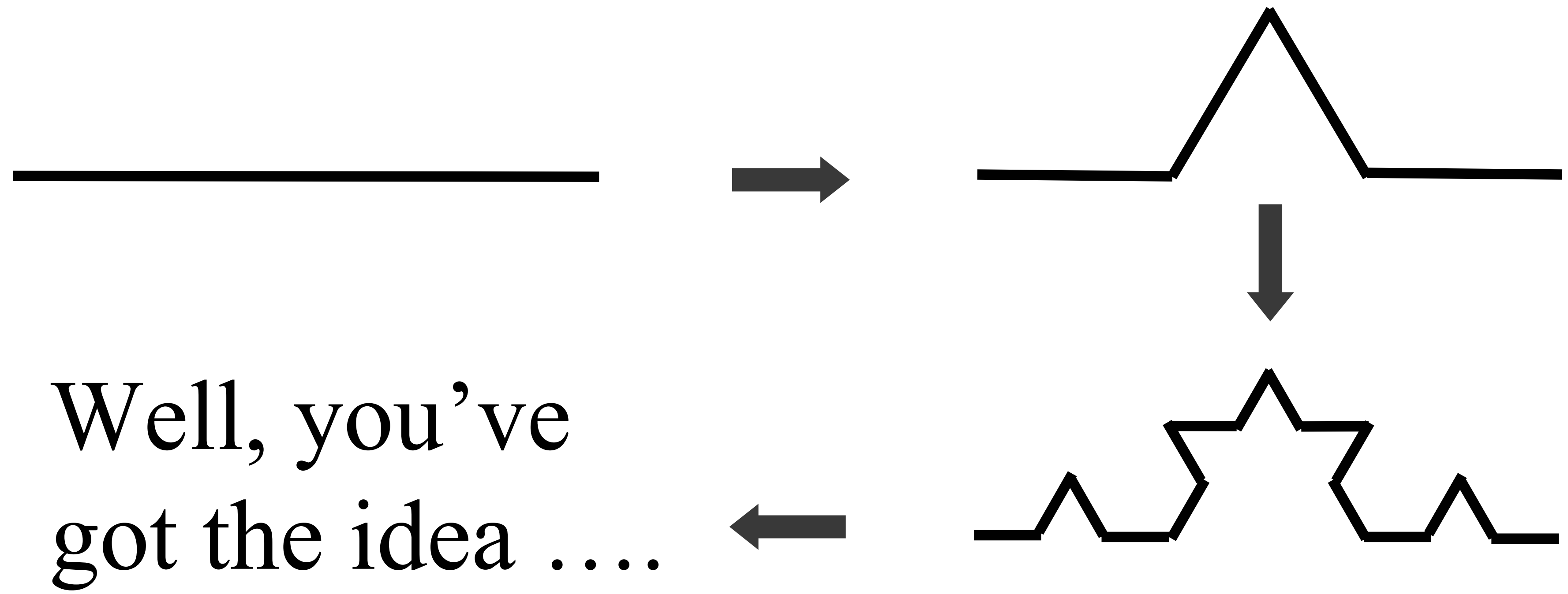
- Exact definition: a fractal is a geometrical object, the Hausdorff dimension of which is strictly larger than its topological dimension.
- Without entering into details, this definition cannot be used in practice because it implies examining all possible covers of the object by sets of finite radius.
- In practice, the estimation of Hausdorff dimension is restircted to the study of covers of the object by balls of various radii.

- For a “normal” curve:



- If radius r is 2 times smaller, the number N of balls is 2 times large: $N \propto r^{-D}$, with $D = 1$

- Koch's snowflake is a fractal obtained iteratively:

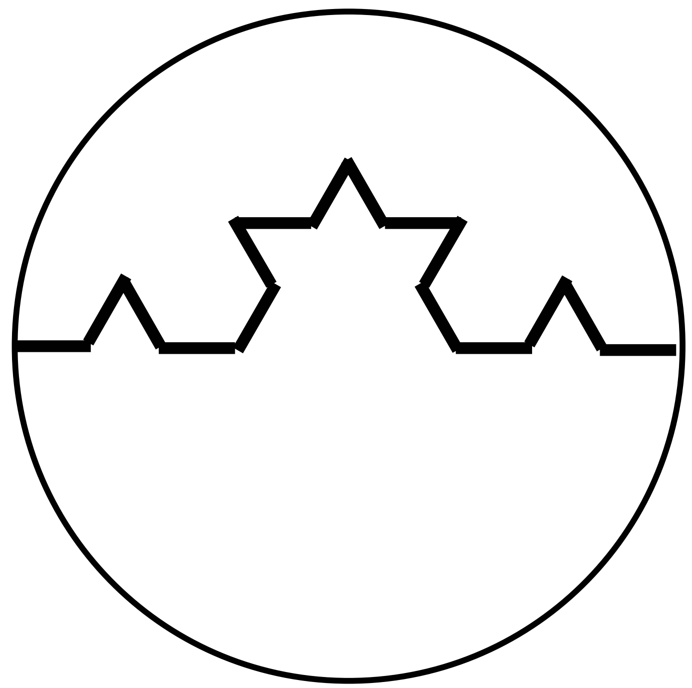


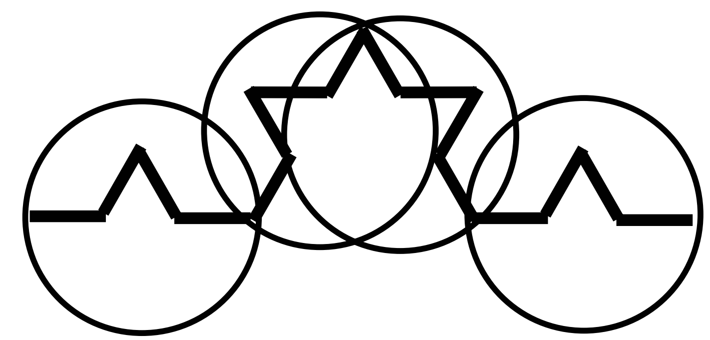
- The limit object is of infinite length (factor 4/3 on the length at each iteration), but it is bounded and has null volume. It is "more than a curve but less than a surface."

FRACTAL DIMENSION (4)

- But if the following cover is used:

$r_1 = 1$
$N_1 = 1$





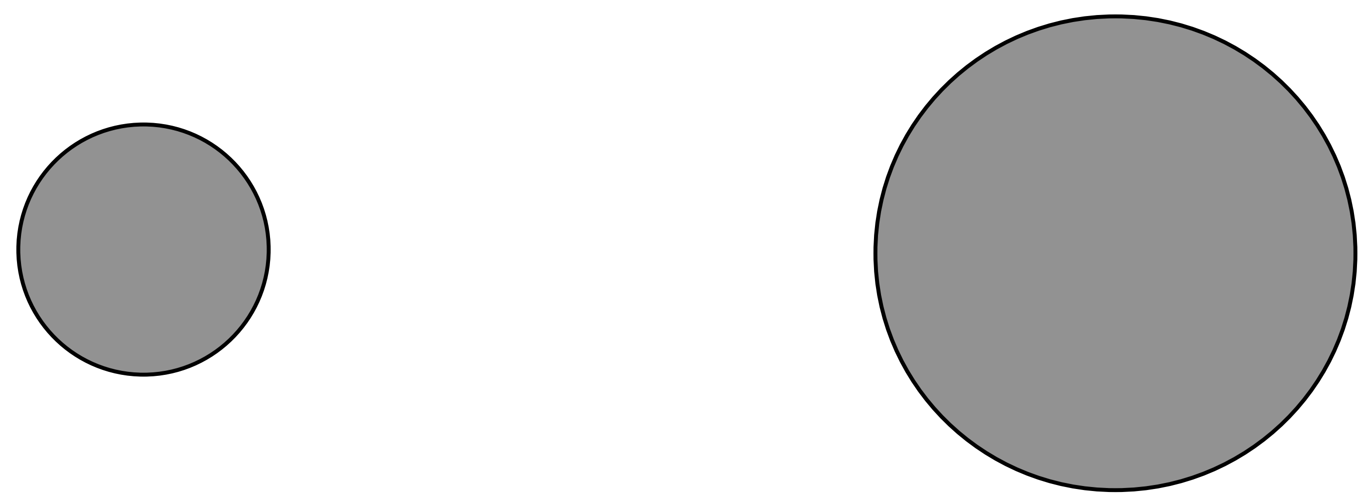
$r_2 = 1/3$
$N_2 = 4$

- Thus, if $N_k = \mathrm{C}\, r_k^{-D}$:

$$\frac{N_2}{N_1} = \left(\frac{r_2}{r_1}\right)^{-D} \quad \Rightarrow \quad D = \log\left(\frac{N_2}{N_1}\right) \Big/ \log\left(\frac{r_1}{r_2}\right) = \frac{\log 4}{\log 3} \approx 1.26$$

- This value indeed indicates that Koch's snowflake is less than a surface, but more than a curve. The fractal dimension quantifies the *occupation of the space* containing the object.
- Fractal objects are characterized by *scale invariance*: if one observes a part of a fractal at a smaller scale, the strcture is the same as for the whole fractal.

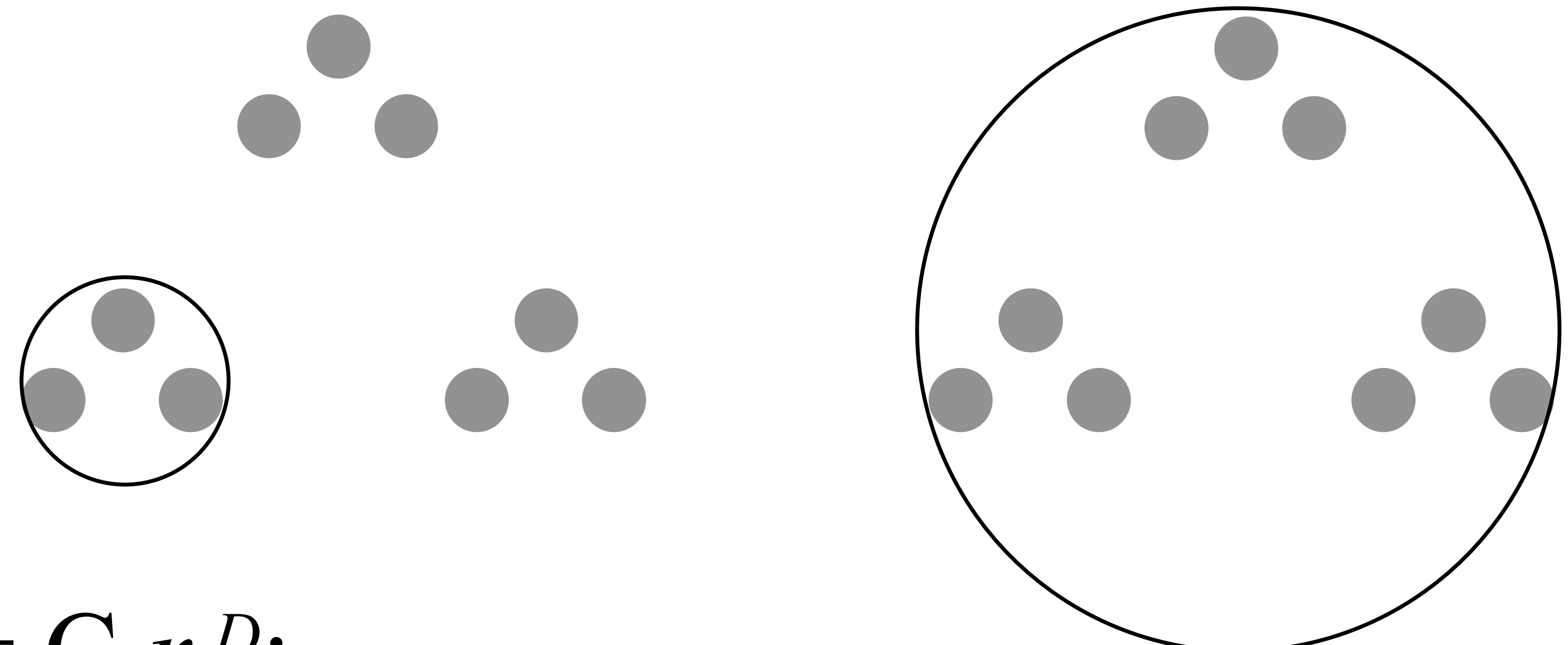
- Equivalently, one may study the evolution of some quantity (≈ mass) with respect to radius. For a homogeneous object:



- If the radius is 2 times larger, the surface is 4 times larger, thus $M \propto r^D$, with $D = 2$.

- For the following fractal object:

$r_1 = 1$
$M_1 = 3$



$r_2 = 4$
$M_2 = 9$

- So, if $M_k = C\, r_k^D$:

$$\frac{M_2}{M_1} = \left(\frac{r_2}{r_1}\right)^D \Rightarrow D = \log\left(\frac{M_2}{M_1}\right) \Big/ \log\left(\frac{r_2}{r_1}\right) = \frac{\log 3}{\log 4} \approx 0.79$$

- The two approaches give the same value for the fractal dimension for “perfect” (obtained iteratively) fractals.
- Of course, for non regular fractals (such as Hénon’s map attractor), they must be obtained through an averaging process.
- By all means, in practical situations, only a finite number of points will be available.

- Lyapunov exponents will be introduced in the continuous time context, but extension to the discrete time case is immediate.

$$\frac{dX(t)}{dt} = G[X(t)] \;\rightarrow\; X(t) = G_t[X(0)]$$

- For a close initial condition:

$$G_t[X(0)+\varepsilon] = G_t[X(0)] + J_t\varepsilon + O(\|\varepsilon^2\|)$$

with J_t the Jacobian: $J_t = \left.\dfrac{\partial G_t(X)}{\partial X}\right|_{X=X(0)}$

- One can show that the limit matrix:

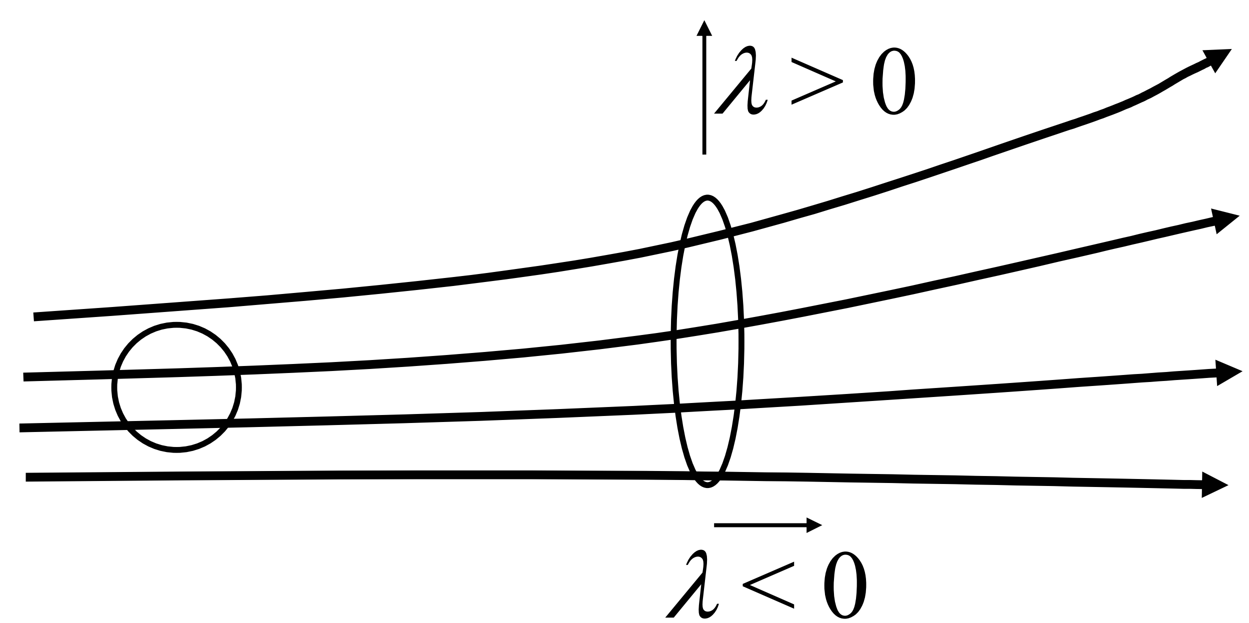
$$\Lambda_{X(0)} = \lim_{t\to\infty} \left[J_t^{\mathsf{T}} J_t \right]^{1/2t}$$

exists and does not depend on $X(0)$

- Logarithms $\{\lambda_i\}$ of the eigenvalues of this matrix are called Lyapunov exponents.
- For an attractor with null volume, one must have:

$$\sum \lambda_i < 0$$

- A chaotic dynamics is characterized by at least one positive Lyapunov exponent.
- Lyapunov exponents quantify the expansion or contraction rates in the eigendirections of flow.



- Most of the time, only one time series is available. How is it possible to estimate the time evolution of state vectors?
- Imbedding theorem:

One can reconstruct the attractor up to a diffeomorphism from a scalar time series $\{x(n)\}$ using the vectors:

$$X(n)=[x(n),x(n+\tau),x(n+2\tau),\cdots,x(n+(m-1)\tau)]^{\mathrm{T}}$$

with $m > 2D$ and τ almost arbitrary. But this suppose an infinite number of noiseless samples.

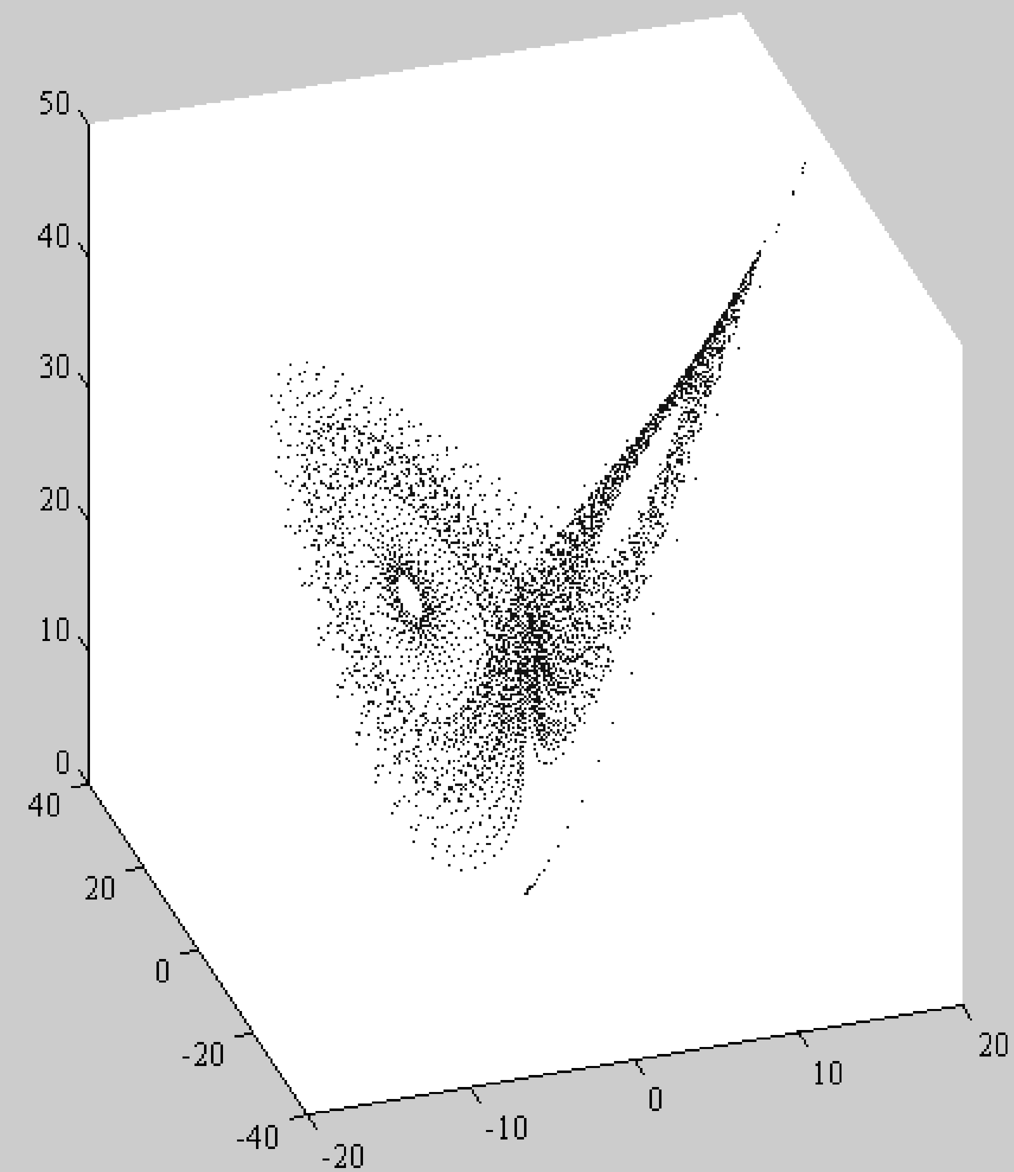
- Equivalence up to a diffeomorphism implies that such as fractal dimension and Lyapunov exponents are not modified.
- With a finite number of samples, one starts by determining an appropriate value for τ, then for the embedding dimension m, since fractal dimension D is of course not known in advance.
- Condition $m > 2D$ can often be slacken to $m > D$.

- Example: reconstruction of Lorenz attractor defined by:

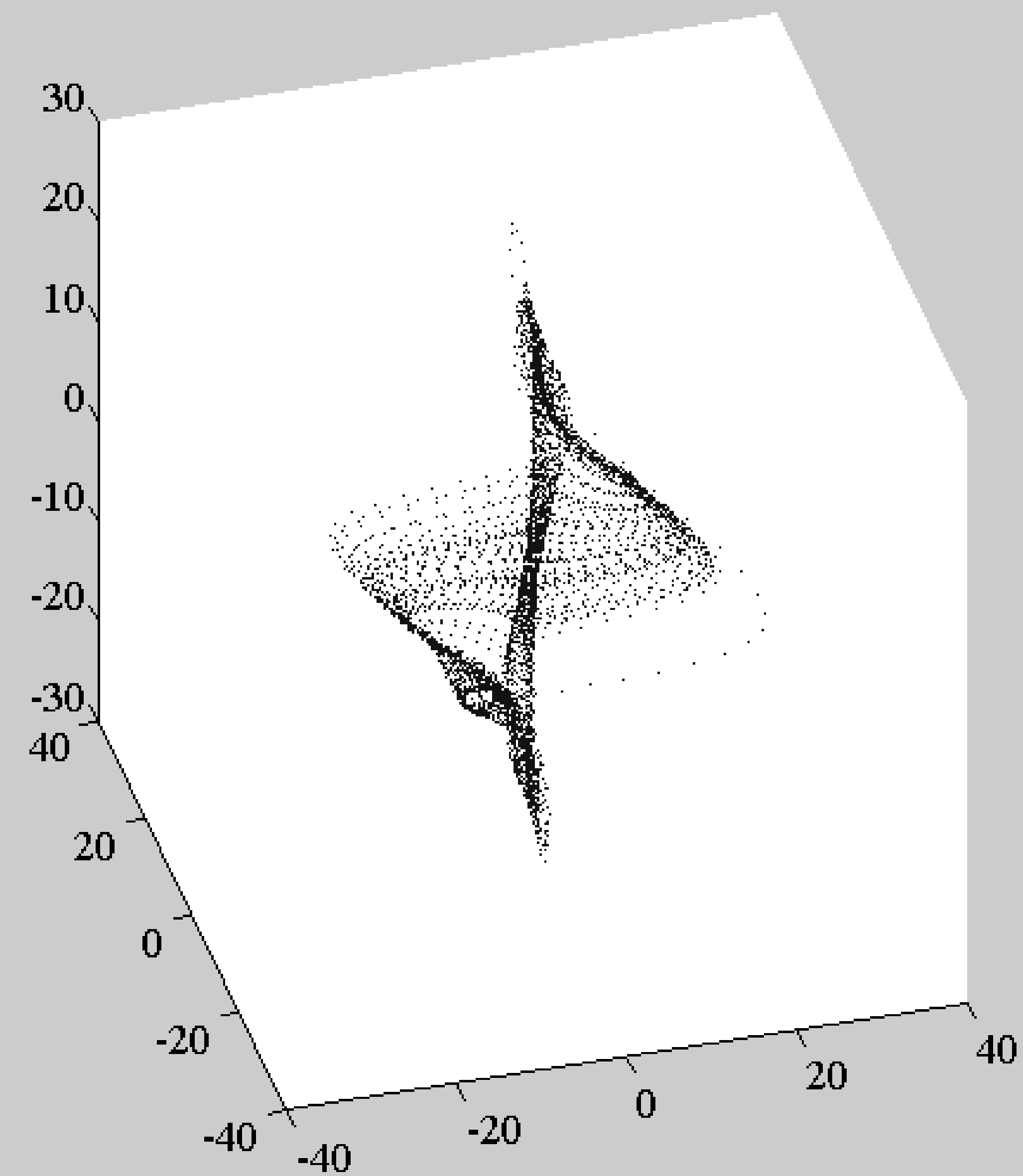
$$\begin{cases} \dfrac{dx(t)}{dt} = 10[y(t) - x(t)] \\ \dfrac{dy(t)}{dt} = x(t)[28 - z(t)] - y(t) \\ \dfrac{dz(t)}{dt} = x(t)y(t) - \dfrac{8}{3}z(t) \end{cases}$$

- The attractor is reconstructed from samples of $y(t)$.

ATTRACTOR RECONSTRUCTION (4)



Lorenz attractor



reconstructed attractor

DETERMINATION OF DELAY τ (1)

- Components of the reconstructed vectors must not be:
  - Too close, because then the reconstructed attractor is on the diagonal.
  - Too far apart (independent), because the structure of the original attractor is lost.
- The first method proposed consisted in choosing τ as the position of the first zero crossing of the autocovariance function of the signal.

But this approach typically gives too large values.

- It was also proposed to select τ as the position of the first minimum of mutual information between samples.
- In practice, a good solution consists in taking τ as the position where the autocovariance is (1-1/e) times its maximum value.

- First method proposed: analysis of the evolution with respect to the embedding dimension m of the effective dimension of the space generated by the vectors of the reconstructed attractor.
- This can be done by computing the SVD of the matrix built by line stacking of the reconstructed vectors, which amounts to compute the eigenvalues of their covariance matrix. Then, a test can be performed on the singular values to extract the effective dimension.

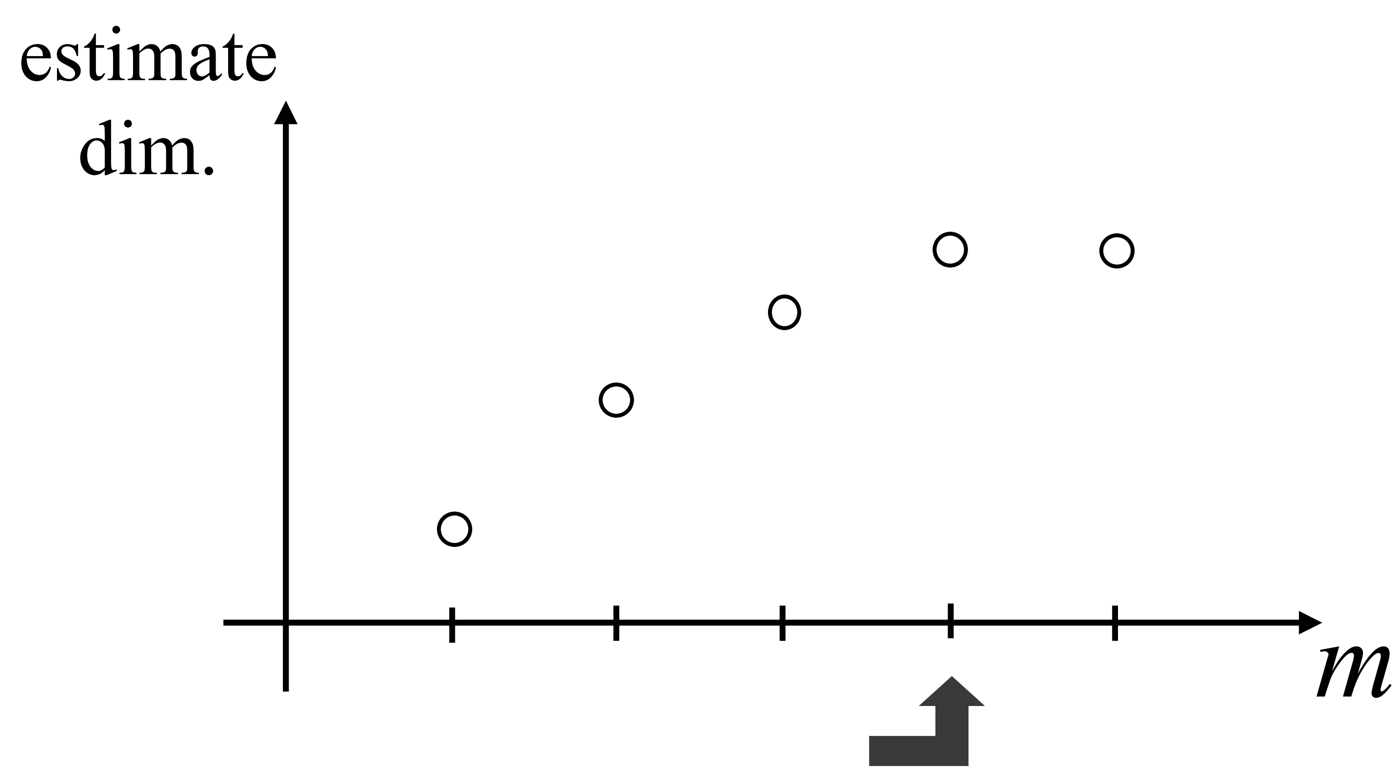
- False neighbor method

One increases the immersion dimension until vectors that were previously neighbors do not separate anymore. For vector $X(k)$ and $X_{nn}(k)$ its nearest neighbor at a distance $d_m(k)$ for dimension m, one measures:

$$E = \frac{\left| x(k+m\tau) - x_{PV}(k+m\tau) \right|}{d_m(k)}$$

If E is above some threshold (typically between 10 and 50), then $X(k)$ et $X_{pv}(k)$ are “false neighbors” for dimension m.

- But a simple and efficient approach consists in applying a method for fractal dimension estimation for increasing values of m and observe when the estimate saturates.



ATTRACTOR DIMENSION ESTIMATION (1)

- Estimation by cover

It is simpler to use a cover by cubes.

- Estimation of point dimension

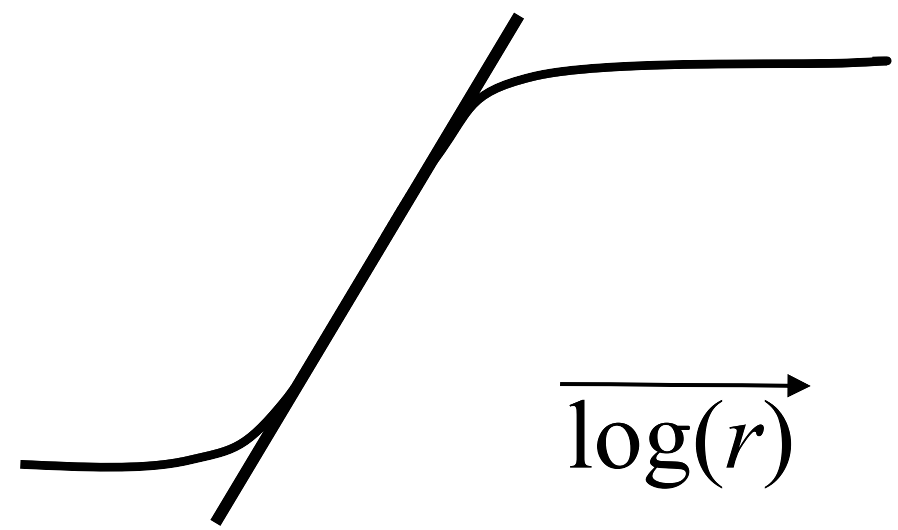
One increases the radius of a sphere centered on a point, and the "mass" computed is the number of points in this sphere for all radii. This is repeated on all points and the evolutions of mass versus radius are averaged.

- Unfortunately these methods are not robust. A more efficient approach, introduced by Grassberger and Procaccia, consists in using for the "mass" the square of point density in a sphere. This corresponds to what is called <u>correlation dimension</u>. One computes:

$$M(r) = \frac{2}{N(N-1)} \sum_{\substack{1<i,j<N \\ i\neq j}} \theta(r - \| X(i) - X(j) \|)$$

with $\theta(u)=0$, $u<0$, $\theta(u)=1$, $u>0$.

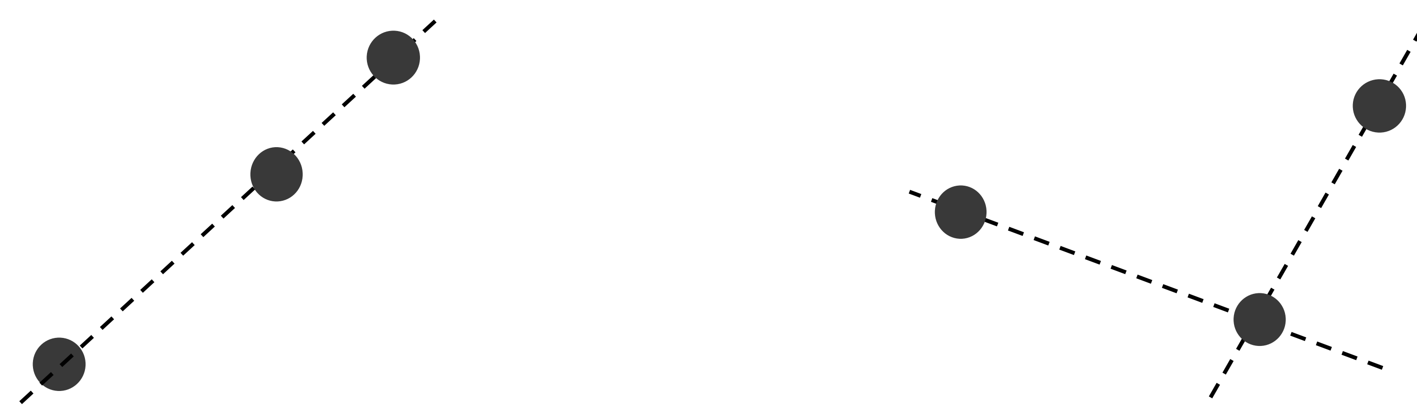
- In practice, if $M(r) = c.r^D$, one estimates the slope of $\log[M(r)]$ with respect to $\log(r)$.



- Of course this must be done in the linear part. When r is too small, there are only few pairs of points closer than r, and when r is too large, all pairs of points are closer than r.

- Local Intrinsic Dimension (LID)

A different approach consists in interpreting the fact that the fractal dimension quantifies the occupation of embedding space by the attractor. For a point and its closest neighbors:



local dimension = 1

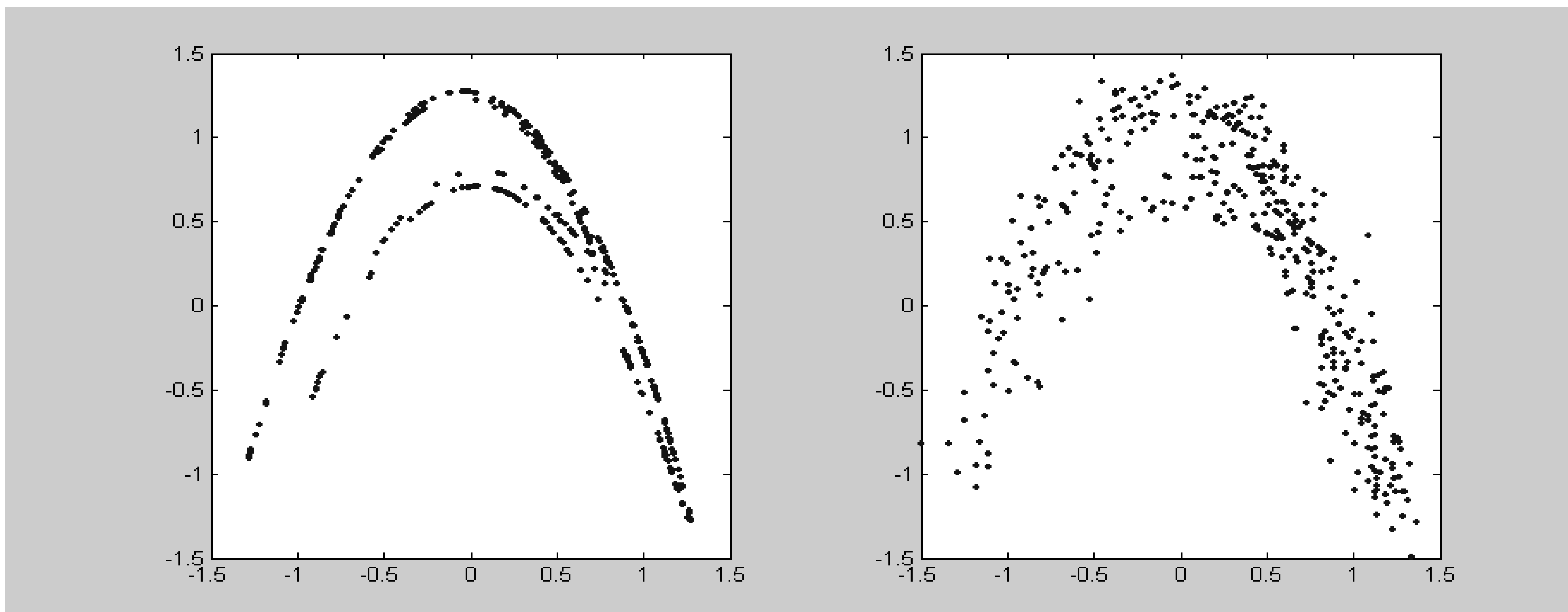
local dimension = 2

- Obviously in practice points will not be perfectly aligned. In fact, on selects randomly a vector X and its k ($k>m$) nearest neighbors $\{X_{(i)}\}$. Then the matrix:

$$A = \lfloor X_{(1)} - X, X_{(2)} - X, \cdots, X_{(k)} - X \rfloor$$

is built and its effective rank is computed using SVD. The process is iterated on a suitable number of randomly chosen vectors and the LID is the average of the effective ranks.

- Unfortunately, the presence of additive noise "blows up" the attractor, which loses its fractal aspect



original attractor attractor + noise (snr 40 dB)

- Estimation of all the exponents

One picks up a vector $X(n)$ at random, and determines its k nearest neighbors $\{X(i_n)\}$. One has:

$$X(i_n+1)-X(n+1)=\delta(i_n+1)\approx \mathbf{J}_n\delta(i_n)$$

The Jacobian $\mathbf{J}_n$ is estimated by minimizing:

$$\sum_{i=1}^{k}\| \delta(i_n+1)-\mathbf{J}_n\delta(i_n)\|^2$$

This operation is repeated on $X(n+1)$, (determination of the k nearest neighbors …), up to an index $n+N$-1.

The exponents are estimated using:

$$\lambda_p = \frac{1}{N} \log(\Lambda_p)$$

with Λ_p the pth eigenvalue of the matrix product $\prod \mathbf{J}_{n+j}$, $j=0, \ldots, N-1$.

- It is necessary in practice to average the results on many trajectories.

- Estimation of the largest exponent

By all means, it is usually the most interesting value, and a robust estimation algorithm has been proposed.

It is based on the fact that the largest exponent λ dictates trajectory separation, with the distance evolving as:

$$\mathrm{d}(t) = c.\exp(\lambda t)$$

One picks at random a vector $X(n)$, and its closest neighbor $X(m)$ is determined. One has:

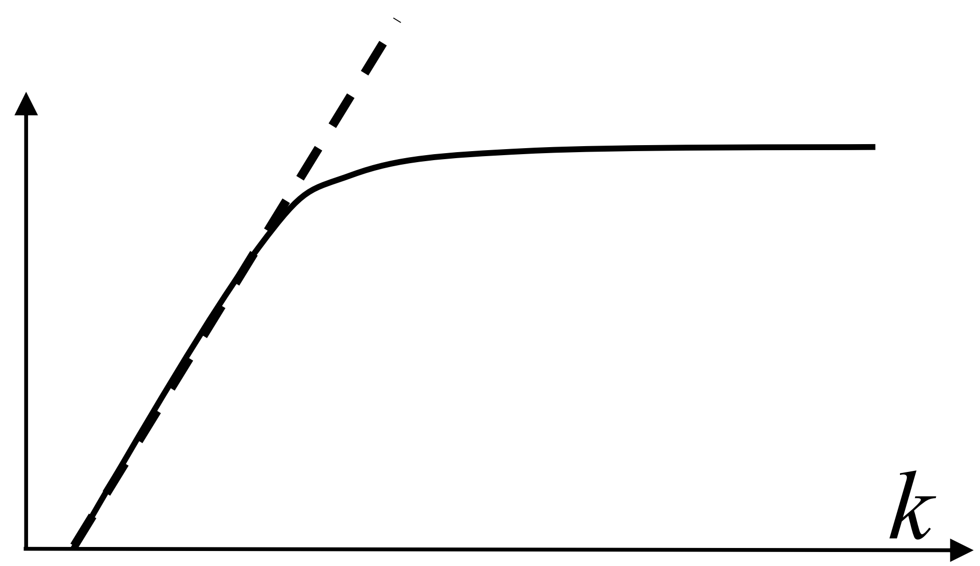
$$d_n(0) = \| X(n) - X(m) \|$$

$$d_n(k) = \| X(n+k) - X(m+k) \| \approx d_n(0) \exp(\lambda k)$$

thus:

$$\log[d_n(k)] \approx \lambda k + \log[d_n(0)]$$

This operation is repeated on a sufficiently large number of randomly chosen vectors, the evolution of log-distances with respect to *k* are averaged, and then the slope is estimated in the linear part:



Saturation of course takes place as soon as the distance between vector pairs is of the order of attractor diameter.

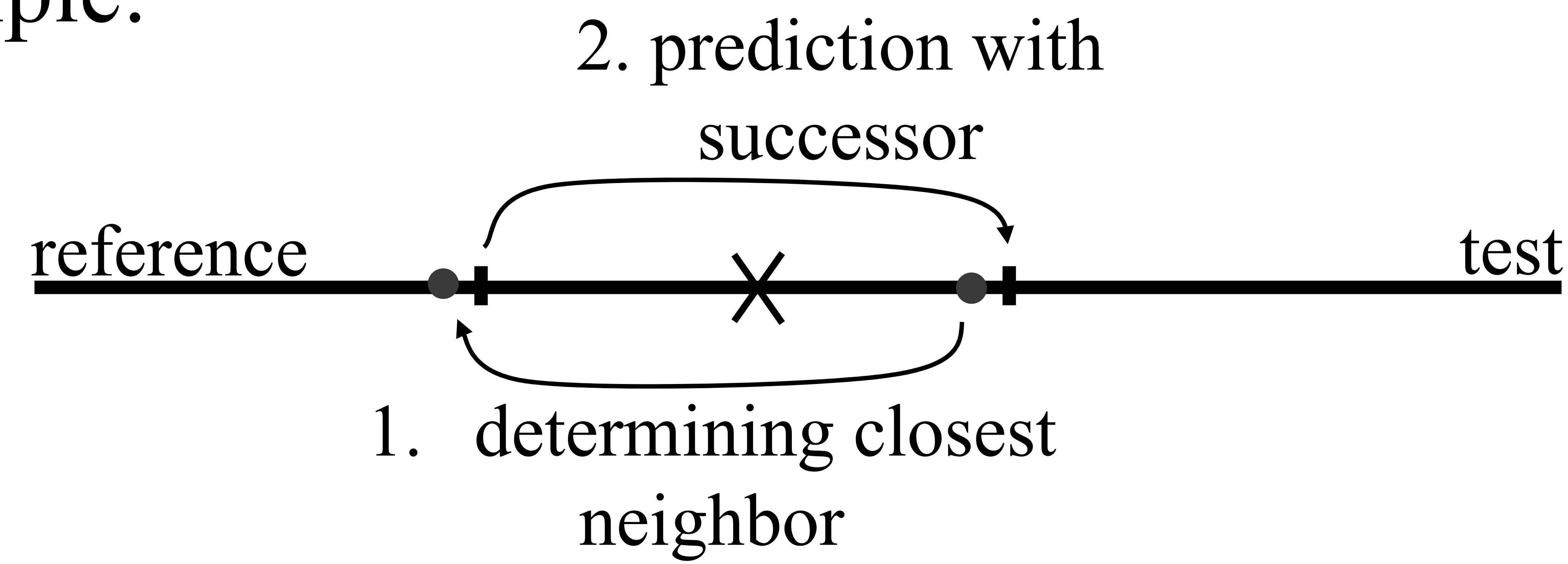
- To estimate attractor dimension D, the number of samples must be in the order 10^D to 40^D.
- To estimate Lyapunov exponents, the number of samples must be larger than 40^D.
- If only the largest exponent is estimated, around 5^D to 10^D samples is enough.
- Note that if D is large and the number of samples is too small, one does not "see" the structure of the attractor.

- This type of prediction, suited to a chaotic dynamics, is beased on the following simple idea:

Of course, a chaotic dynamics implies an exponentially fast separation of trajectories. But this dynamics is deterministic, and on the short term, *to close vectors will correspond close successors*.

- Thus, if two vectors $X(n)$ et $X(p)$ are close, the first components $x(n+m\tau)$ and $x(p+m\tau)$ of their successor will be close too.
- To test if a dynamics can be predicted efficiently in this way, one splits the samples into two groups (which gives the same partition for the reconstructed vectors).
- The test part is used to assess prediction performance, the reference part to find neighboring vectors.

- Principle:

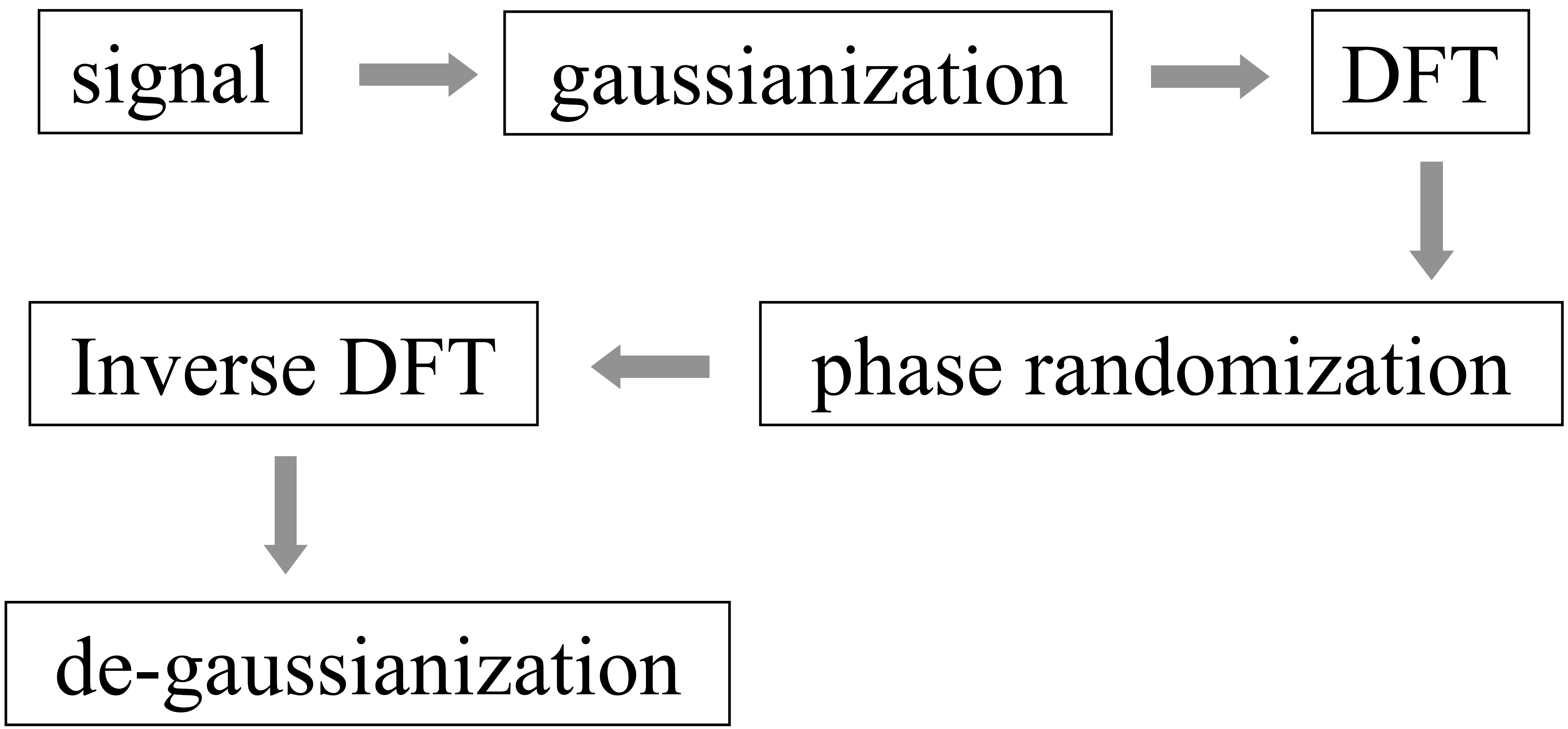


- One can also use several neighbors, and define the prediction as a sum of successors weighted by the inverses of the distances.

SURROGATE DATA (1)

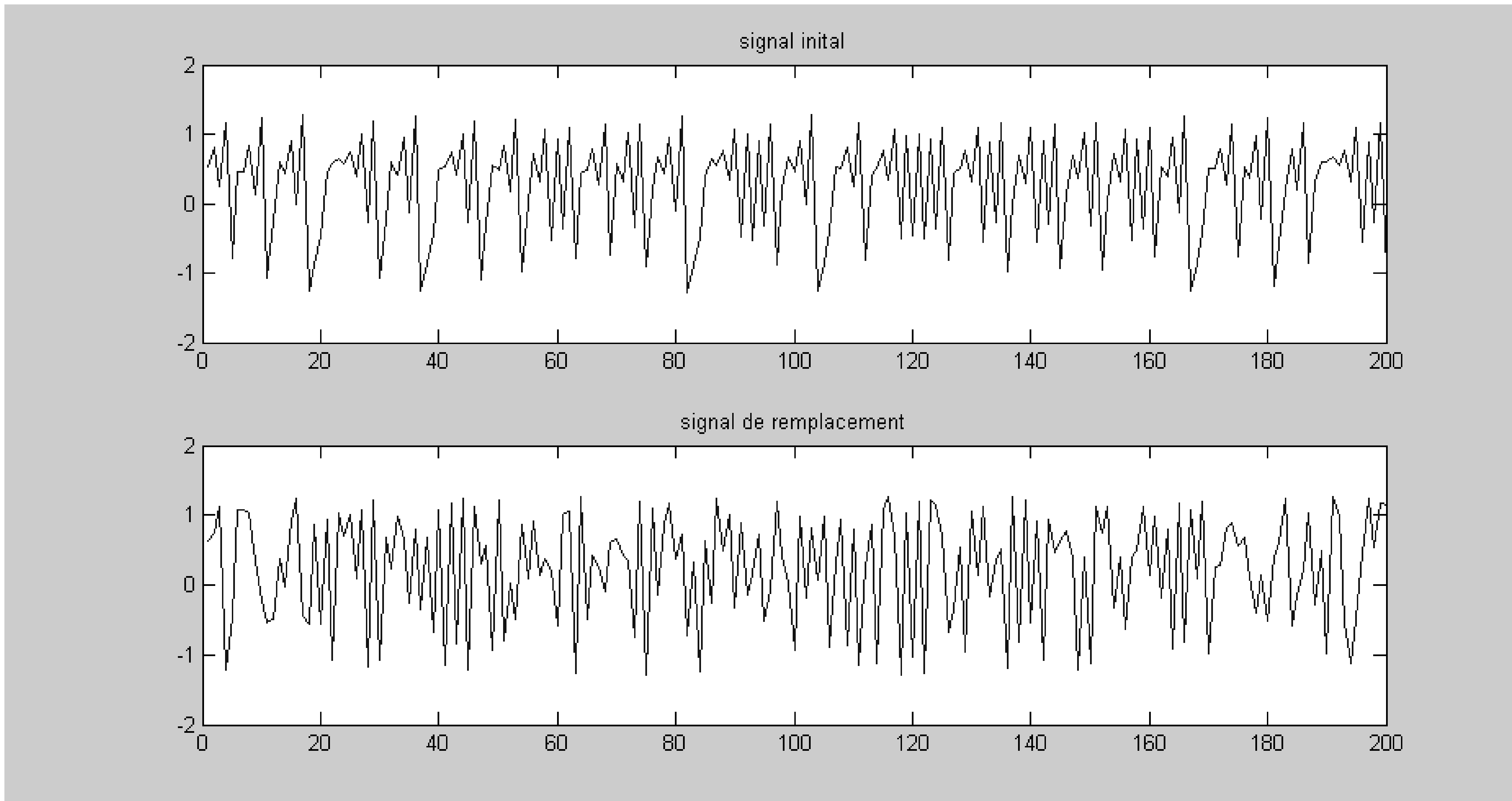
- Surrogate data can be used to:
  - Test the presence of nonlinear dynamics
  - Test the significance level of the characteristics (fractal dimension, Lyapunov exponents, predictibility…) obtained.
- To build these surrogates, one uses the fact that linear relationships between samples imply only 2nd-order statistics, i.e. the autocorrelation function, which is even and doest not carry any phase information.

- Principle of surrogate generation:

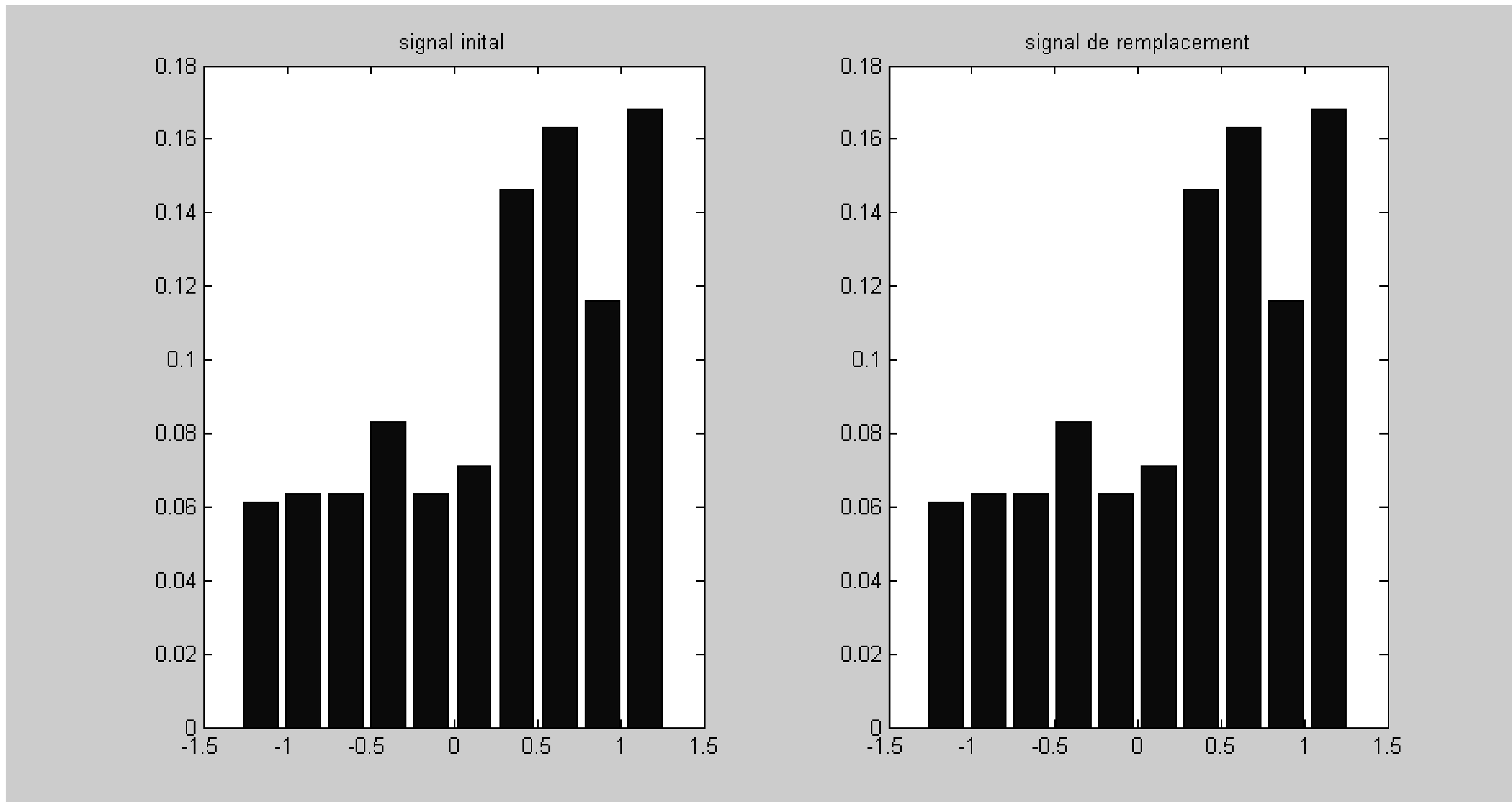


- To “Gaussianize“ the samples, one feeds them through an instantaneous nonlinearity which is the distribution of the samples.
- Phase randomization on the discrete Fourier transform (phases uniformly drawn between 0 and 2π), destroys any potential nonlinear structure.
- De-Gaussianization consist in applying the inverse of the instantaneous linear transform.

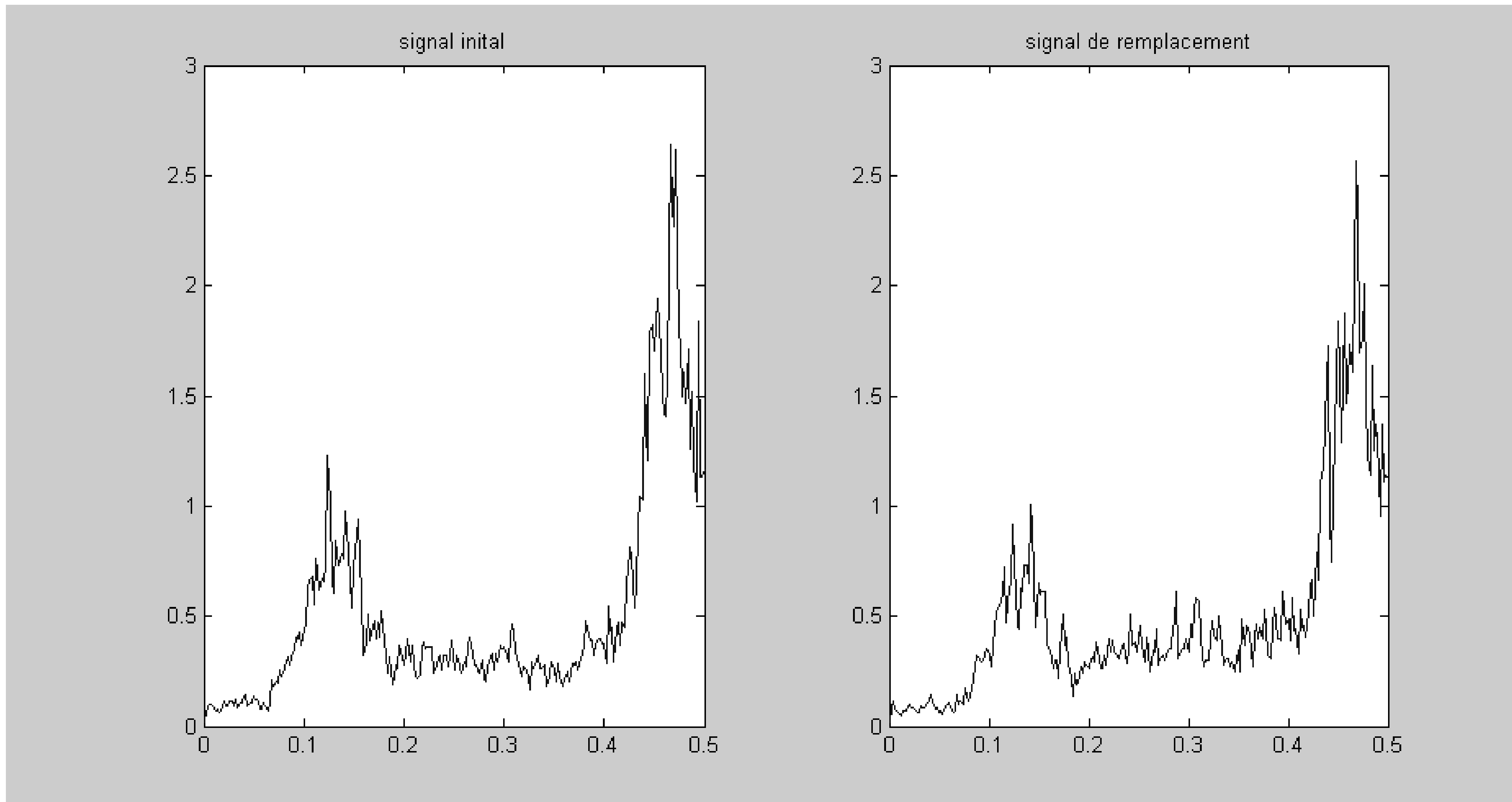
- Example: surrogate signal for Hénon



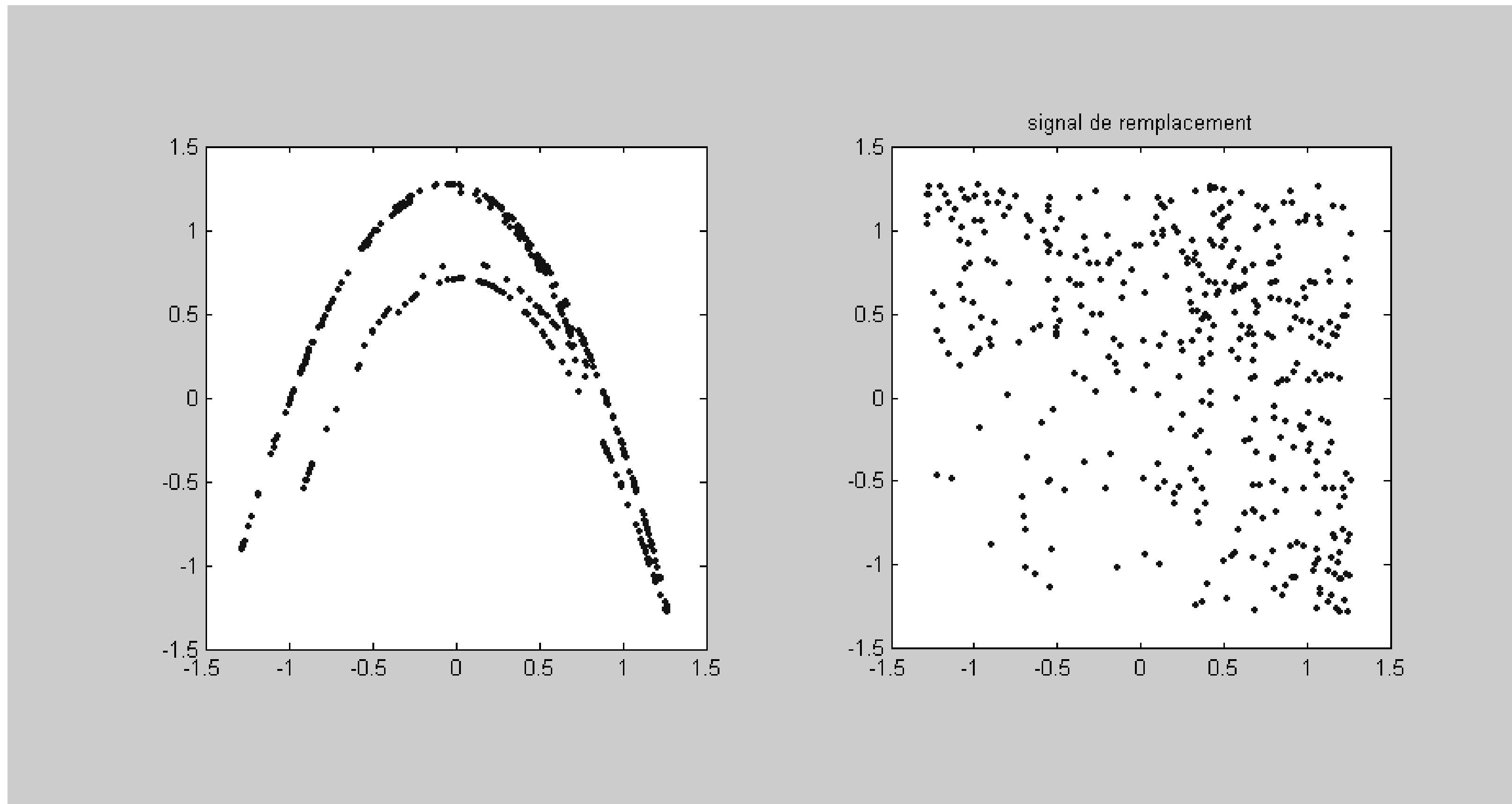
Estimated probability density functions:



Estimated power spectra



But for the attractors…



- No surprise: the chaotic dynamics is responsible for attractor structure. If it is suppressed, then the structure disappears.

REFERENCES

1. H.D.I. Abarbanel et al., ″Analysis of observed chaotic data in physical systems,″ *Rev. Mod. Phys.*,vol. 65, no. 4, 1993, pp. 1331-1391.
2. J. Argyris, G. Faust, and M. Haase, *An Exploration of Chaos*, North Holland, 1994.
3. T Gautama, D. P. Mandic, and M. M. van Hulle, ″A novel method for determining the nature of time series,″ *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, May2004, pp. 728-736.