Dimensions, Fractals, and Dimensions from Time Series
Fundamentals of Analyzing Biomedical Signals

Dimensions

1-cube  2-cube  3-cube  4-cube
5-cube  6-cube  7-cube  8-cube
9-cube  10-cube  11-cube  12-cube

Ioannis Keppler: Harmonices mundi, 1619

from: autodeskresearch.com
Euclidean geometry

- characterization of some geometric object
- integer dimension

<table>
<thead>
<tr>
<th>object</th>
<th>dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>point</td>
<td>0</td>
</tr>
<tr>
<td>line</td>
<td>1</td>
</tr>
<tr>
<td>area</td>
<td>2</td>
</tr>
<tr>
<td>volume</td>
<td>3</td>
</tr>
<tr>
<td>$n$-cube</td>
<td>$n$</td>
</tr>
</tbody>
</table>

- number of degrees of freedom for characterization

*Time series analysis*: minimum number of equations needed to model a physical system, system complexity, number of degrees of freedom (see later)
Euclidean geometry and generalized dimension

Idea:

If you multiply all lengths by $a$,

- lengths will change by a factor $a$
- areas will change by a factor $a^2$
- volumes will change by a factor $a^3$
- ...

$\rightarrow$ determine dimension by exponent of content-scaling
Non-Euclidean geometry

> generalized concept (F. Hausdorff, 1919)
> dimension of some non-Euclidean object in an $m$-dim. space
> idea:

- cover object in $m$-dim. space with hypercubes of side length $\varepsilon$
- determine minimum number $N(\varepsilon)$ of hypercubes necessary to fully cover object
- we find:

$$N(\varepsilon) \propto \varepsilon^{-D_0}$$

- Hausdorff dimension, fractal dimension, box-counting dimension ($D_0 = D_H$ in most cases)
Hausdorff dimension of a line

\[ D_0 = \frac{\log\left[\frac{N(\varepsilon)}{N(\varepsilon')}\right]}{\log(\varepsilon / \varepsilon')} = \frac{\log 3}{\log 3} = 1 \]
Hausdorff dimension of Cantor set
Hausdorff dimension of Cantor set

\[ D_0 = \frac{\log \left[ \frac{N(\varepsilon)}{N(\varepsilon')} \right]}{\log (\varepsilon / \varepsilon')} = \frac{\log 2}{\log 3} = 0.6309 \]
Fractals

**Def.:** a set $\mathfrak{F}$ is called a *fractal*, if

- $\mathfrak{F}$ has some fine structure
- $\mathfrak{F}$ is irregular
- $\mathfrak{F}$ shows self-similarity (a subset of $\mathfrak{F}$ is similar to $\mathfrak{F}$)
- fractal (Hausdorff-Besicovitch) dimension strictly exceeds the topological dimension

**Applications:**

many natural structures, modelling, technology, art, …
Hausdorff and topological dimension of Cantor set

\[ D_0 = \frac{\log \left[ \frac{N(\varepsilon)}{N(\varepsilon')} \right]}{\log(\varepsilon / \varepsilon')} = \frac{\log 2}{\log 3} = 0.6309 \]

Length \( L \) (topological dimension)

\[ L = 1 - \frac{1}{3} - \frac{2}{9} - \frac{4}{27} \ldots = 1 - \frac{1}{3} \sum_{v=0}^{\infty} \left( \frac{2}{3} \right)^v = 0 \]

\[ D_0 > L \]
Fractals

Q: What does the “B.” in “Benoît B. Mandelbrot” stand for?

A: “Benoît B. Mandelbrot”

possibly not a joke
Fractals and the Coastline Paradox

If the coastline of Great Britain is measured with a ruler of 100 km length, then the length of the coastline is about 2800 km. With a ruler of 50 km length, the total length is about 3400 km, i.e. 600 km longer. For rulers with smaller length, the length of the coastline diverges to infinity.
Some Fractals and their dimensions

- Cantor set \((\log_3(2)=0.6309)\)
- dendrite of Julia set \((1.2)\)
- Koch curve \((\log_3(4)=1.2619)\)
- boundary of dragon curve \((1.5236)\)
- Sierpinski triangle \((1.5849)\)
- Hexaflake \((\log_3(7)=1.7712)\)
- Sierpinski carpet \((\log_3(8)=1.8928)\)
- fractal pyramid \((\log_2(5)=2.3219)\)
- Menger sponge \((\log_3(20)=2.7268)\)
Some Natural Fractals

- Romanesco broccoli
- Lung
- Blood vessels mouse brain
- River delta
- Fern
- Dandelion
- Bolt
Some Natural Fractals

>> Physics News Update, 92, 19 August 1992
The landscape of DNA may be fractal
(Phys Rev Lett 22 Jun 92)

Fractal patterns inside cells can reveal breast cancer
(Phys Rev Lett, 12 Jan 98)

>> Physics News Update, 399, 26 Oktober 1998
Tumor growth can be fractal (fractal dimension: 1.21)
(Phys Rev Lett, 2 Nov 98)
Technical Fractals (an example)

fractal grids to generate turbulent flow

towards “fractal wind turbine blades”

Fractals and Art
Fractals and Art

Jack the dripper: chaos in modern art

Of all the abstract expressionist painters, Jackson Pollock was perhaps the most controversial. He would dash around large canvases rolled out on the floor of his barn, dripping paint from a wooden stick. The critics poured scorn on his paintings, calling them "meaningless chaos". But chaos is now a rigorous scientific concept that we know appears throughout nature. One important part of chaos theory is fractal behaviour, which describes objects that have similar patterns when viewed at different magnifications. Richard Taylor, a physicist at the University of New South Wales, has now discovered this characteristic in many of Pollock's works. Rather than being the fraud that many people assume, Taylor believes that Pollock subconsciously understood the patterns of nature so well that he was able to capture their very essence -- chaos and fractals -- on canvas.

Physics Web November 1997
Fractals and Art
Fractals and Art

Fractals determine date of paintings

[4 Jun 1999] Paintings by the late Jackson Pollock - considered to be one of the fathers of modern art - can be dated by fractal geometry according to Australian physicists (Nature 399 422). Pollock's artwork during the late 1940s consisted of paint dripped from a can onto large canvases spread out on the floor of his barn. Richard Taylor, Adam Micolich and David Jonas from the University of New South Wales in Sydney discovered that the fractal dimension of Pollock's drip paintings increased from nearly 1.0 in 1943, to 1.72 in 1952, suggesting that Pollock gradually refined his technique over time to make his painting more fine grained.
Strange attractors and fractal dimensions

- Ikeda map (1.7)
- Zaslavskii map (1.39 ?)
- Hènon map (1.261)
- Rössler system (2.01)
- Lorenz system (2.06)
Generalized dimensions
- estimating dimensions in high-dimensional space via box-counting is hard
- box counting ignores how densely the boxes are populated
- idea: weight boxes by probability $p_i$ to find state in box $i$

Rényi dimensions, q-dimensions
- partition state space into $M$ hypercubes (boxes) of side length $\varepsilon$
- estimate probability by $p_i = \lim_{N \to \infty} \frac{N_i}{N}$

\[ D_q := \lim_{\varepsilon \to 0} \frac{\log \left( \sum_{i=1}^{M(\varepsilon)} p_i^q \right)}{(q-1) \log(\varepsilon)} \]
Generalized dimensions

- $D_0 := \lim_{q \to 0} D_q$ (box-counting dimension)
- $D_1 := \lim_{q \to 1} D_q$ (information dimension)
- $D_2$ (correlation dimension)
- $D_{topo} \leq D_H \leq \ldots \leq D_2 \leq D_1 \leq D_0 \leq m$
- in most cases: $D_H = \ldots = D_2 = D_1 = D_0$

$D_0$ counts non-empty boxes
$D_1$ measures gain of information to find state in box $i$
if $D_0 = D_1$, attractor is homogeneous
Generalized dimensions

- $D \notin \mathbb{N}$: strong indicator for nonlinearity (chaotic dynamics)
  
  ($D$ diverges for purely stochastic dynamics)

- characterizes self-similarity, complexity

- provides hints for modelling (degrees of freedom, attractor structure)

- sanity check via embedding theorems
Generalized dimensions

dimension from time series:

we have:

\[ D_q := \lim_{\epsilon \to 0} \frac{\log \left( \sum_{i=1}^{M(\epsilon)} p_i^q \right)}{(q-1) \log(\epsilon)} \]

- replace limes by slope in double-logarithmic plot
- approximate \( p_i \approx N_i / N \)
- approximate sum over probabilities by “correlation sum”:

\[ C_q(\epsilon) := \frac{1}{N} \sum_i \left( \frac{1}{N} \sum_j \Theta (\epsilon - |\vec{v}_i - \vec{v}_j|) \right)^{q-1} \]

counts number of point closer than \( \epsilon \)

Generalized dimensions

dimension from time series:
correlation dimension:

\[ D_2 := \lim_{\epsilon \to 0} \frac{\log \left( \sum_{i=1}^{M(\epsilon)} p_i^2 \right)}{\log(\epsilon)} \approx \lim_{\epsilon \to 0} \frac{\log(C_2(\epsilon))}{\log(\epsilon)} \]

• quickest to calculate

Generalized dimensions

dimension from time series:

Rényi dimensions

element: sine wave

from: G. Ansmann
Generalized dimensions

dimension from time series:

Rényi dimensions

dimension from time series: sine wave
Generalized dimensions

dimension from time series:

Rényi dimensions

example: Gaussian noise

from: G. Ansmann
Generalized dimensions

dimension from time series:

\[ D_2 \sim m \]

from: G. Ansmann
Generalized dimensions

dimension from time series:

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; b = 0.3 \]

Rényi dimensions

element: Hénon map

from: G. Ansmann
Generalized dimensions
dimension from time series:

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; b = 0.3 \]

Rényi dimensions
example: Hénon map

from: G. Ansmann
Generalized dimensions

dimension from time series:

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; b = 0.3 \]
Generalized dimensions

dimension from time series:

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; b = 0.3 \]

from: G. Ansmann
Generalized dimensions

dimension from time series:

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; \ b = 0.3 \]
Generalized dimensions

dimension from time series:

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; b = 0.3 \]

from: G. Ansmann
Generalized dimensions

dimension from time series:

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; b = 0.3 \]

from: G. Ansmann
Generalized dimensions

dimension from time series:

\[
\begin{align*}
    x_{n+1} &= 1 - ax_n^2 + y_n \\
    y_{n+1} &= bx_n
\end{align*}
\]

where

\[a = 1.4; b = 0.3\]

literature:

\[D_2 \sim 1.26\]

Rényi dimensions

dimensional example: Hénon map

from: G. Ansmann
Generalized dimensions

dimension from time series:

dimension depends on the scope

How many dimensional is a plate of spaghetti? Zero when seen from a long distance, two on the scale of the plate, one on the scale of the individual noodles and three inside a noodle.

Maccaroni is even worse.

attributed to Peter Grassberger
Generalized dimensions

dimension from time series:
simulating macaroni: sum of two incommensurable sine waves (tube/torus) and some noise (dough):

from: G. Ansmann
Generalized dimensions

dimension from time series:

Rényi dimensions

example: torus

\[ C_2 \]

from: G. Ansmann
Generalized dimensions

dimension from time series:

Rényi dimensions

example: torus

from: G. Ansmann
Generalized dimensions
dimension from time series:

\[ D_2 \]

Rényi dimensions
equation: torus

from: G. Ansmann
Generalized dimensions

dimension from time series:

field applications

- number of data points \((\lim N \to \infty)\)
- data precision \((\lim \epsilon \to 0)\)
- strong correlations in data (sampling interval)
- noise
- filtering
- superposition of non-interacting dynamical systems
Generalized dimensions

dimension from time series:

number of data points

- requirement: \( \lim N \to \infty \)
- field applications: \( N \) always limited, stationarity issues, system life-time, observation time

- proposed estimators:
  \( N \sim 10^{D_2} \) (Albano et al., 1987)
  \( N \sim 42^{D_2} \) (Smith, 1988)
  \( N \sim 100^{D_2} \) (Procaccia, 1989)

- \( N \) as large as possible; resolvability of attractor structure depends on density of phase space points
Generalized dimensions

dimension from time series:
number of data points determines maximum resolvable dimension
(Ruelle criterion)

Generalized dimensions from time series:

Along the lines of the ‘science and fiction’ title of this talk, let me conclude on a lighter note. Readers of ‘The hitchhiker’s guide to the galaxy’, that masterpiece of British literature by D. Adams, know that a huge supercomputer has answered ‘the great problem of life, the universe, and everything’. The answer obtained after many years of computation is 42. Unfortunately, one does not know to what precise question this is the answer, and what to make of it. It think that what happened is this. The supercomputer took a very long time series describing all it knew about ‘life, the universe, and everything’ and proceeded to compute the correlation dimension of the corresponding dynamics, using the Grassberger–Procaccia algorithm. This time series had a length $N$ somewhat larger than $10^{21}$. And you can imagine what happened. After many years of computation the answer came: dimension is approximately $2 \log_{10} N \approx 42$.

Generalized dimensions

dimension from time series:

data precision

- field applications: analog-digital converter (ADC with $n$ bits)
- digitizing accuracy: $p = A/2^n$, $A =$ amplitude range
- quantization error: $q = p/2$

Rényi dimensions

what can go wrong?

Fig. 1. Slopes of log-log plots of the correlation integral (estimates of $D_2$) calculated for data rounded to different resolutions, for (a) the Hénon attractor, (b) the Roessler attractor. In (a) one scaling region and in (b) three slightly different scaling regions were evaluated (digitizing accuracy $p$, average length scale $\epsilon$).

The values of $\Delta$ are calculated by a least squares fit of the slope of the log $C(\epsilon)$ versus log $\epsilon$ plots over a selected scaling region $\epsilon_{\text{min}} \leq \epsilon \leq \epsilon_{\text{max}}$ (represented by an average length scale $\epsilon = (\epsilon_{\text{min}} \epsilon_{\text{max}})^{1/2}$). The data can be well fitted by the equation

$$\Delta = D_2(1 - kp/\epsilon),$$

(2)

where $p$ is one half of the least significant digit and $k$ is a positive factor of order unity.

Fundamentals of Analyzing Biomedical Signals

**Dimensions**

**Generalized dimensions**
**dimension from time series:**

- data precision

possible way to minimize influence of the quantization error:
- add noise prior to digitizing
  (pre-whitening, dithering, bleaching)

effectiveness depends on system under study
- not effective for broad-band signals

**Rényi dimensions**
**what can go wrong?**

caveat: adding noise can lead to erroneous dimension estimates

---

Generalized dimensions

dimension from time series:

- strong correlations in data

field applications

- sampling rate according to Nyquist-Shannon theorem:

  \[
  \text{at least twice as high as signal's maximum frequency } f_{\text{max}}
  \]

  to avoid aliasing

- how to treat cases with unknown \( f_{\text{max}} \)?
- how to treat a chaotic signal?
- is resampling (over-/undersampling) a good choice?
Generalized dimensions

dimension from time series:

strong correlations in data

problem: for a sufficiently fine temporal resolution, points close in time are also close in phase space

→ correlation sum overestimated

**Theiler correction:**

exclude temporally close points from the correlation sum:

$$\sum_{i,j} \Theta (\epsilon - |\vec{v}_i - \vec{v}_j|) \rightarrow \sum_{|i-j|>W} \Theta (\epsilon - |\vec{v}_i - \vec{v}_j|)$$

(adjust normalization accordingly)

Generalized dimensions

dimension from time series:

strong correlations in data

how to choose cutoff $W$ for Theiler correction?

minimum requirement: $W$ in the order of autocorrelation time ($\Delta$)

better: $W > \Delta \left( \frac{2}{N} \right)^{\frac{2}{m}}$

(exact choice is then insignificant)

Fundamentals of Analyzing Biomedical Signals

Dimensions

Generalized dimensions

dimension from time series:

strong correlations in data

(ex.: Lorenz system without correction)

Rényi dimensions

what can go wrong?

from: G. Ansmann
Fundamentals of Analyzing Biomedical Signals

Dimensions

**Generalized dimensions**

**dimension from time series:**

strong correlations in data

(ex.: Lorenz system with correction)

**Rényi dimensions**

what can go wrong?

\[ D_2 \]

from: G. Ansmann
Generalized dimensions

dimension from time series:

noise

field applications:
- data always noisy (characteristics of noise?)
- measurement errors (white noise approximation)
- additive vs. multiplicative noise
- if number of data points limited, dimension of white noise finite!

Rényi dimensions

what can go wrong?
Generalized dimensions
dimension from time series:
example: white noise; $N = 8192$

\[ D_2^{\text{max}} = 2 \log_{10}(8192) \approx 7.8 \]
Generalized dimensions
dimension from time series:
example: low-pass-filtered noise; $N = 8192$, Theiler correction

\[ D_2 \sim m \]

Rényi dimensions
what can go wrong?

---

Generalized dimensions

dimension from time series:

example: low-pass-filtered noise; $N = 8192$, Theiler correction

Rényi dimensions

what can go wrong?
Generalized dimensions

dimension from time series:

noise

classical filtering of noise induces structure in phase space
→ spatial (long-ranged) correlations

Theiler correction only minimized short-ranged correlations

⇒
- do not use classical filter for chaotic signals!
- apply other methods to discriminate determinism from stochasticity
- use other nonlinear noise reduction schemes (future lectures)
Generalized dimensions

dimension from time series:

filtering

field applications:
- sampling theorem, avoid aliasing
- noise reduction (see above!)
- chaotic signals typically broad-band (see Linear Methods)
- do not filter chaotic signals!

Rényi dimensions

what can go wrong?
Generalized dimensions

dimension from time series:

example: filtered Hénon map

\[
\begin{align*}
x_{n+1} &= 1 - ax_n^2 + y_n \\
y_{n+1} &= bx_n \\
z_{n+1} &= \exp(-\eta)z_n + x_n
\end{align*}
\]

(a=1.4; b=0.3)

recursive realization of single-pole low-pass filter (1. order)

\[
\dot{z} = -\eta z + x
\]

\[\text{FIG. 3. Influence of degree of filtering on attractor shape. Shown are return maps (Z_{n+1} vs Z_n) of the “filtered Hénon system,” Eq. (4), with eight-bit resolution. All four attractors are scaled to the same size.}\]

Generalized dimensions

dimension from time series:

eExample: filtered Hénon map

dimension increase (+1)

interpretation:
- superposition of two systems (Hénon system + filter)
- filter (passive, linear)
  one-dimensional system

filtering does not affect other invariant measures

Rényi dimensions

what can go wrong?

FIG. 1. $D_2$ values for simulated low-pass filtering with different rolloff frequencies $\eta$ for the Hénon map. The solid line shows the predicted $D_1(\eta)$. Note change of scale at the dashed lines.

Generalized dimensions

dimension from time series:

product dynamical system

- Consider two non-interaction dynamical systems $A$ and $B$. For the product dynamical system $A \times B$, we have

$$\dim(A \times B) \leq \dim(A) + \dim(B).$$

- Consider time series of system observables $v(A)$, resp. $v(B)$ that solely depend on $A$, resp. $B$, and a time series $v(A \times B)$ of the product dynamical system with

$$v(A \times B) = \alpha \, v(A) + \beta \, v(B)$$

- Consider cases $\alpha = \beta$, $\alpha < \beta$, and $\alpha > \beta$
Dimensions

Generalized dimensions

dimension from time series:

product dynamical system

Rényi dimensions

what can go wrong?
Generalized dimensions

dimension from time series:

product dynamical system

sine wave (A) + Hénon map (B)

\[ \alpha = \beta = 1 \]

\[ \alpha = 0.1; \beta = 1 \]

\[ \alpha = 1; \beta = 0.1 \]
Dimensions

Generalized dimensions

Rényi dimensions

dimension from time series:

$\alpha = \beta = 1$
Fundamentals of Analyzing Biomedical Signals

Generalized dimensions

dimension from time series:

\[ \alpha = 0.1; \ \beta = 1 \]

Rényi dimensions

what can go wrong?
Generalized dimensions

dimension from time series:

\[ \alpha = 1; \beta = 0.1 \]
Generalized dimensions

dimension from time series:

product dynamical system

sine wave (A) + Hénon map (B)

Rényi dimensions

what can go wrong?

\[ \alpha = \beta = 1 \]

\[ \alpha = 0.1; \beta = 1 \]

\[ \alpha = 1; \beta = 0.1 \]
Estimate dimension of time series via slope of correlation sum

- check multiple embedding dimensions $m$
- select scaling region properly
- apply Theiler correction
- be aware of influencing factors, limitations, and pitfalls

$D \notin \mathbb{N}$: chaotic dynamics

$D \in \mathbb{N}$: hint at regular dynamics

$D = \infty$: noise