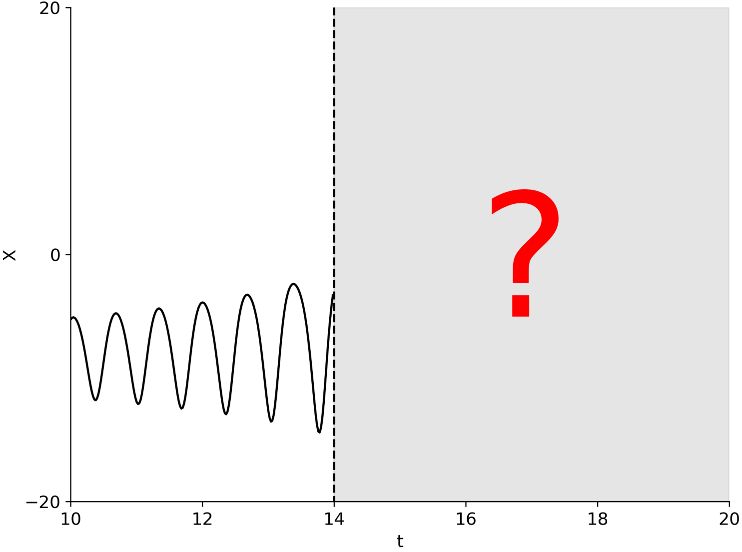
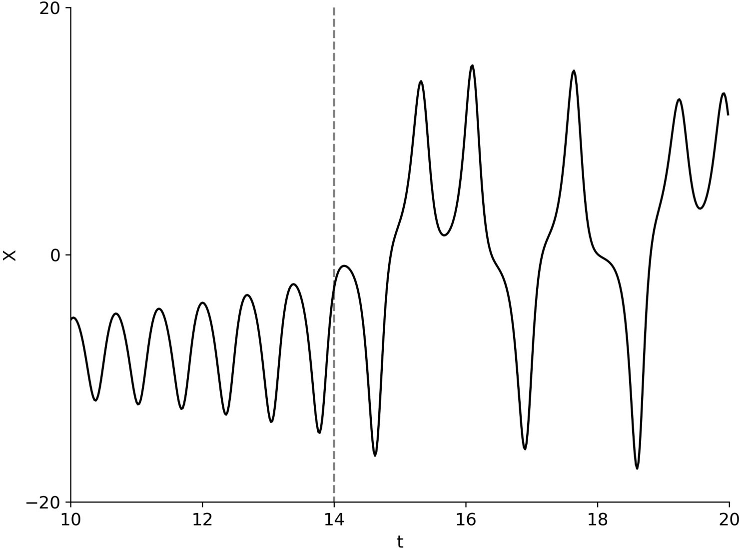
Learning Chaos

Mark Vousden
University of Southampton

George Frederic Watts – Chaos c. 1875


20
0
−20
X
10
12
14
16
18
20
t
?


20
0
−20
X
10
12
14
16
18
20
t

Learning Chaos

Mark Vousden
University of Southampton

Before the ocean and the earth appeared– before the skies had overspread them all–
the face of Nature in a vast expanse was naught but Chaos uniformly waste.
It was a rude and undeveloped mass, that nothing made except a ponderous weight;
and all discordant elements confused, were there congested in a shapeless heap.

(Ovid, Metamorphoses)

Chaos

"When the present determines the future but the approximate present does not approximately determine the future."

An attempt at definition:

Deterministic: The future is entirely determined by the past.

Sensitive: A small perturbation drastically affects the future.

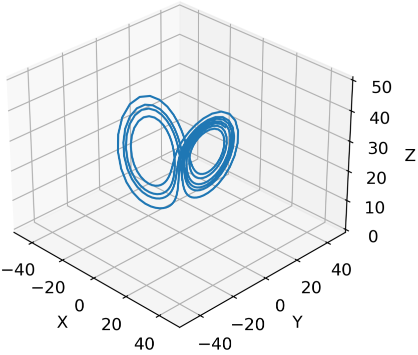
Aperiodic: No "end-points" or "cycles".

Butterflies

The Lorenz Attractor (Lorenz '63).

$$dx/dt = \sigma(y - x),$$
$$dy/dt = \rho x - xz - y,$$
$$dz/dt = xy - \beta z.$$

Typically $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$.



Demonstration of Lorenz Attractor Chaos

- Deterministic
- Sensitive
- Aperiodic

Mark plays `lorenz_anim.mp4`.

Chaos is Pervasive

- Meteorology: Met Office public weather service value est. 1.5B GBP (2015). Private work?
- Economics: Predictability of the health of financial systems (SONNETS). Risk management (governments) vs. trader beneficiaries?
- Fluid mechanics
- Computational demography
- Circuits (Chua, PRNG)
- City planning
- Cryptography
- Optimisation/Design Search

Well, we had better get started then.

What do you mean, “started”?!

- Chaos theory:
  - Maxwell’s butterflies (1860s)
  - Poincaré’s three-body problem (1880s)
  - ENIAC predicts the weather (1950)
- Machine learning:
  - McCulloch/Pitts develop first artificial neuron (1943)
  - Turing’s learning machine (1950)
  - Machines rival humans playing Backgammon (1992)

Science has been doing this for $\approx$70 years – why continue?

What do you mean, “started”?!

Science has been doing this for $\approx$70 years – why continue?

In my opinion:

- Rise of GPGPU and increased computational power
- Algorithm advancement – deep learning, ensemble methods
- The big data boom
- Motivation: A deliberate use of chaos in engineering
- Interdisciplinary collaboration

Why Learning?

- Prediction of higher-order behaviour breaks down with system evolution – can the computer do better?
- Defining a numerical model is hard, let alone a computational one. Can we reason purely from measured data?
- Universal approximation theorem suggests feed-forward learning is generally viable, iff (Hornik '90)
  - We throw enough neurons at the problem
  - Our activation function is "sufficiently rich in nonlinearity" (e.g. $\tanh$).

Learning the Lorenz Attractor

One hidden layer, which computes $\mathbf{x} = (x, y, z)^T$ for the next timestep ($\mathbf{x}_{n+1}$) given the present ($\mathbf{x}_n$):

$$\mathbf{x}_{n+1} = \mathbf{W}_2 g(\mathbf{W}_1 \mathbf{x}_n + \mathbf{b}_1) + \mathbf{b}_2$$

Train it on the Lorenz Attractor

$$\begin{aligned} dx/dt &= \sigma(y - x), \\ dy/dt &= \rho x - xz - y, \\ dz/dt &= xy - \beta z, \end{aligned}$$

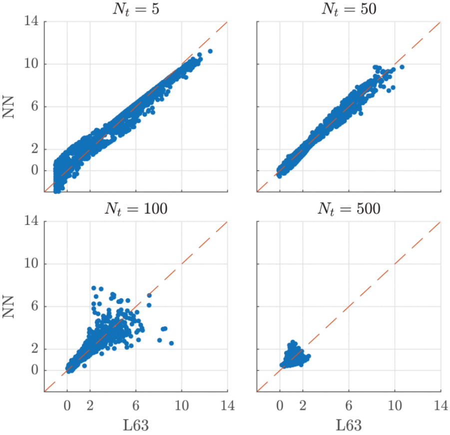
solved by (implicit) Adams Moulton ($dt = 0.01$).

An Elementary Problem

The smallest “viable” case, with a four-neuron hidden layer. Mark plays `lorenz_nn_anim.mp4`.

An Elementary Problem

Not enough neurons/training?



In the end, the sensitivity catches up with you.

$$|\delta\mathbf{x}(t)| \approx e^{\lambda t}|\delta\mathbf{x}_0|$$

Lyapunov exponent $\lambda \approx 0.906$

(Figure: Li '23)
(λ: Viswanath '98)

Why Not Learning?

You can book out Archer for ten years, but:

- You're not learning the chaotic system - you're learning a numerical equivalent (Adams Moulton!)
- Practical measurement of quantities is often difficult.
- Practical systems can have $\gg 10^9$ degrees of freedom.
- Identifying intervention strategies is even harder (particularly if you are only reasoning from data).

In the end, sensitivity overcomes your approximations.

You can't beat the Lyapunov exponent.

So, the problem is hard.


Excellent.

Let's do what any scientist would do...

...and move the goalposts:

- Explore/quantify sensitivity using a chaotic RNN (e.g. an echo state network), training with noisy data (Mahata '23)
- Understand mechanisms, instead of predicting outcomes, with symbolic regression powered by a DNN (Boddupalli '23) $\rightarrow$ intervention analysis?

Summary and Thoughts

- Chaos and machine learning are old, saturated fields of study.
- In a naive approach, you can make your network as complex as you like, but inaccuracies will catch up with you:
  - Integrator (top-down)
  - Data measurement/frequency (bottom-up)
- Better to move the goalposts:
  - What is the sensitivity of my system? To each degree of freedom?
  - How can my intervention alter its behaviour?
  - How big does my computer need to be? How many Lyapunov times is enough?

References, Abridged

Lorenz ’63: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Hornik ’90: 10.1016/0893-6080(91)90009-T

Viswanath ’98: https://www.proquest.com/docview/304430201

Boddupalli ’23: 10.1063/5.0134464

Li ’23: 10.1109/TNNLS.2021.3087497

Mahata ’23: 10.1103/PhysRevE.108.064209