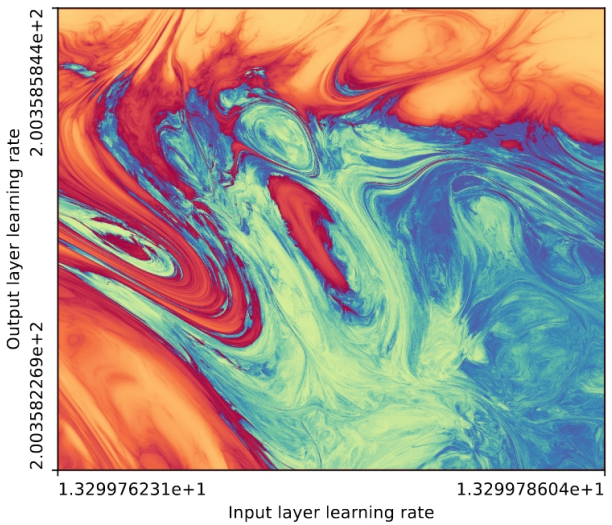

Romanesco broccoli

[cs.LG] 9 Feb 2024

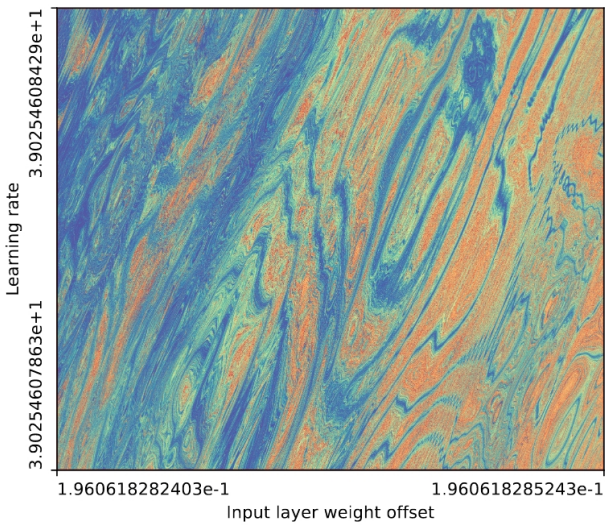
The boundary of neural network trainability is fractal

Jascha Sohl-Dickstein
jascha.sohldickstein@gmail.com

tanh full batch (fractal dim 1.66)



Parameter initialization (fractal dim 1.98)



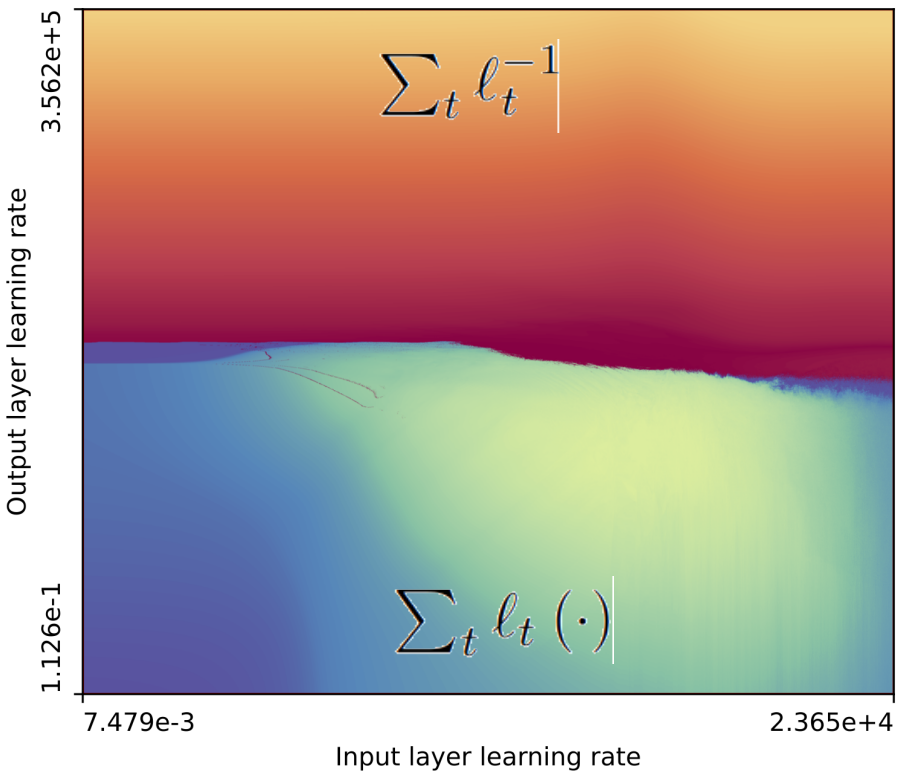
Take home message

- Fractals are beautiful and can happen in any dynamical systems (including optimizing neural networks).
- The boundary of neural network trainability is fractal.

Hyperparameter landscape

- A neural network with one hidden layer of width: n=16
- Training data-set size is equal to the number of model parameters (n^2+n).
- Hyperparameters are the learning rates of input and output layers.
- Weights, data-points, and labels are all randomly initialized from a standard normal distribution,
- Every pixel is one run. (different random seeds make different fractals)
- Red-yellow colors mean training diverged.
  Blue-green colors mean training converged.
- The paler the color the faster the convergence or divergence

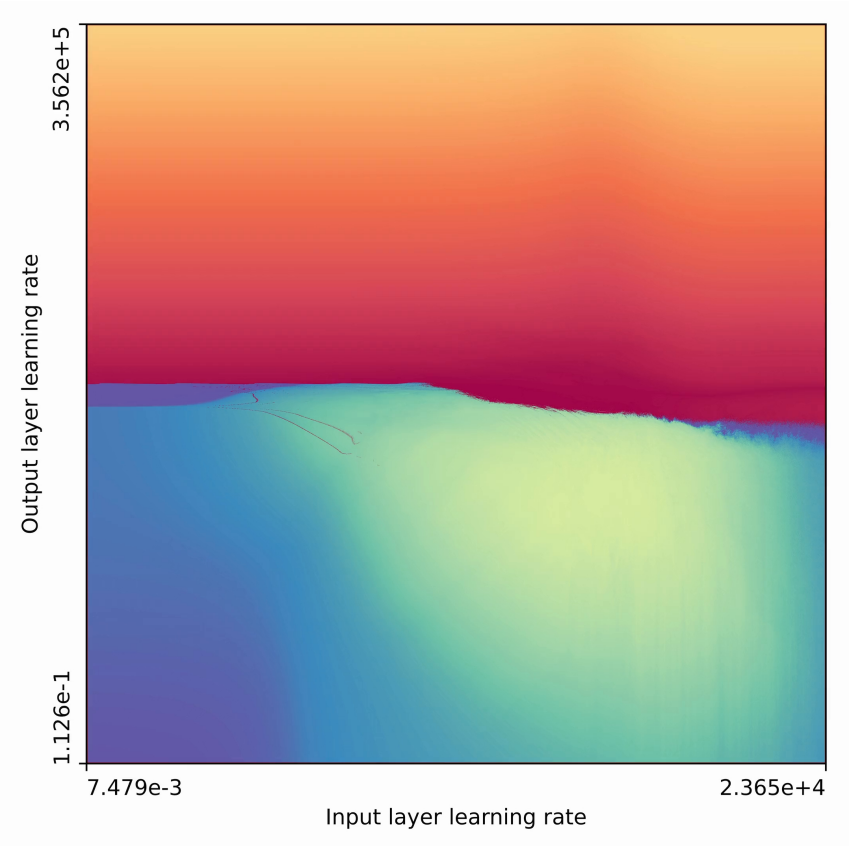
✔ "Best hyperparameters for neural network training are usually very near the edge of stability"



ℓ_t is the loss at training step t

The boundary of neural network trainability is fractal

As we zoom in, we find intricate structure at every scale.

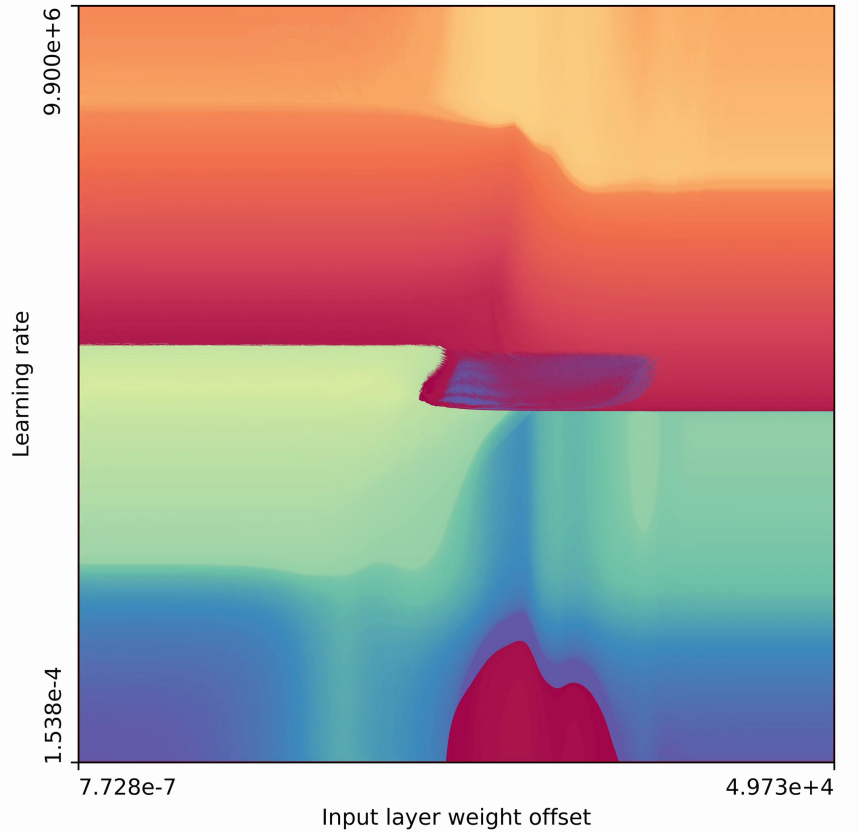


https://player.vimeo.com/video/903855670?h=ca2b077023

The boundary of neural network trainability is fractal

As we zoom in, we find intricate structure at every scale.

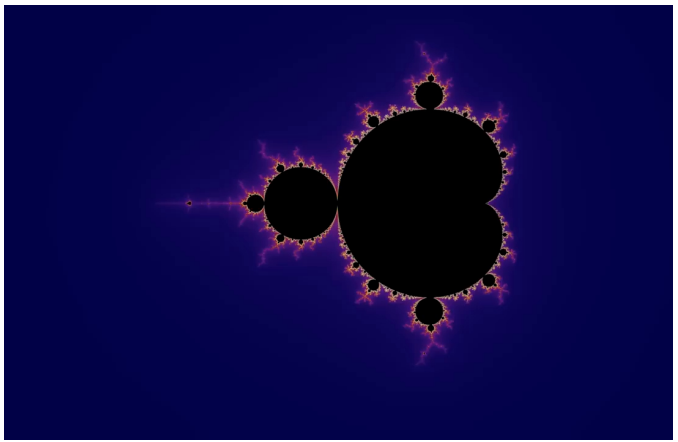
We see fractals if we change the data, change the architecture, or change the hyperparameters.



https://player.vimeo.com/video/903855723?h=ed7eb562bb

Fractals

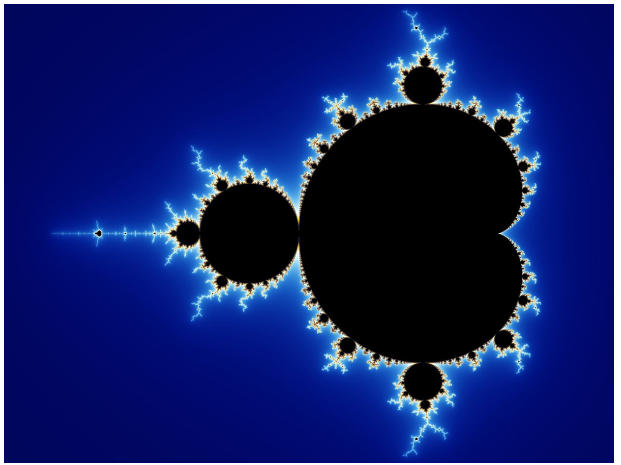
- A **fractal** is a geometric shape that contains **detailed structure at arbitrarily small scales**.
- Many fractals appear **similar at various scales**.
- Fractals are part of a grander subject called **dynamics** that deals with systems that evolve in time. Whether the system in question settles down to equilibrium (**converges**), keeps repeating in cycles, or does something more complicated (e.g, **diverges**).
- The invention of the computer in the 1950s was a turning point in studing fractals and dynamics.



Zooming into the boundary of the Mandelbrot set

https://en.wikipedia.org/wiki/Fractal#/media/File:Mandelbrot_sequence_new.gifMandelbrot

Mandelbrot fractal



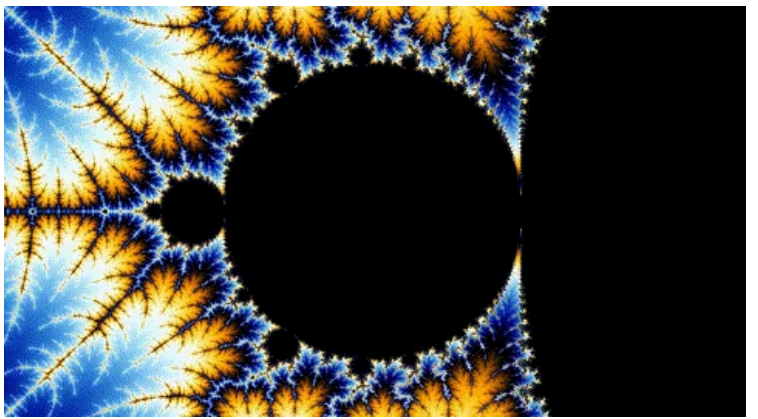
- Set of hyperparameter c for which this iterated function:

$$z_{i+1} = z_i^2 + c$$

diverges or remains bounded (for an initial z value of 0);

- surprisingly complex behavior arises from a very simple mathematical relationship

- It is self-similar under magnification in specific regions



https://upload.wikimedia.org/wikipedia/commons/b/b8/Self-Similarity-Zoom.gif

Take home message

- Fractals are beautiful and can happen in any dynamical systems (including optimizing neural networks).
- The boundary of neural network trainability is fractal.
- Tuning hyperparameters near the trainability boundary can be like dancing with the Devil.
- Read more about dynamics, and develop a dynamical view of the world.
(some books are suggested in the references.)

References

- https://x.com/jaschasd/status/1756930242965606582?s=20
- https://sohl-dickstein.github.io/2024/02/12/fractal.html
- https://arxiv.org/abs/2402.06184
- Gleick, J. (1996). Chaos. London, England: Vintage.
- Strogatz, S. H. (2019). Nonlinear dynamics and chaos (2nd ed.). London, England: CRC Press.

Some amazing fractals:

1. https://www.youtube.com/watch?v=BTiZD7p_oTc&ab_channel=Robindesvilles
2. https://www.youtube.com/watch?v=LhOSM6uCWxk&ab_channel=MathsTown
3. https://paulbourke.net/fractals/magnet/
4. https://en.wikipedia.org/wiki/Lyapunov_fractal