A SMOOTH APPROXIMATION ON THE EDGE OF CHAOS

P. J. POTTS
Department of Computing, Imperial College, London SW7 2BZ.

It is known that for almost all starting points the non-deterministic dynamical system corresponding to a hyperbolic iterated function system with probabilities generates orbits whose frequency of visits to any given Borel subset of the domain is described by the unique invariant measure of the iterated function system with probabilities. In this paper, we will show that under certain conditions this chaotic probability measure can be approximated by a smooth probability density. We apply this result to forgetful neural networks.

1 Introduction

The study of iterated function systems has been an active area of research since the seminal work of Mandelbrot [11] on fractals and self-similarity in nature.

1.1 Iterated Function System

An *iterated function system* (IFS) $\{X; f_1, f_2, \ldots, f_N\}$ on a topological space X is given by a finite set of continuous maps $f_i : X \to X$ $(i = 1, 2, \ldots, N)$. If X is a complete metric space [15] and the maps f_i are all contracting, then the IFS is said to be *hyperbolic*.

A hyperbolic IFS induces a contracting map [10] on the complete metric space $\mathbf{H}X$ of all non-empty compact subsets of X with a unique fixed point called the *attractor* of the IFS.

For graphical applications, X is usually the plane $\mathbb{R}^2$, f_i are contracting *affine* transformations and the attractor is a *fractal*.

1.2 IFS with Probabilities

A hyperbolic IFS with probabilities $\{X; f_1, f_2, \ldots, f_N; p_1, p_2, \ldots, p_N\}$ is a hyperbolic IFS $\{X; f_1, f_2, \ldots, f_N\}$, with X a compact metric space, such that each f_i $(i = 1, 2, \ldots, N)$ is assigned a probability p_i with $0 < p_i < 1$ and $\sum_{i=1}^{N} p_i = 1$.

A hyperbolic IFS with probabilities induces a contracting map [3] on the set of normalized Borel measures $\mathbf{M}X$ on X with a unique fixed point described as a *multifractal* whose support is the attractor of $\{X; f_1, f_2, \ldots, f_N\}$.

In the extreme case where all the maps f_i $(i = 1, 2, \ldots, N)$ are the same, the unique invariant measure has the Dirac delta function [5] as its density and its support is a single point corresponding to the unique fixed point of the map.

In this paper, we show that for a certain class of hyperbolic IFS with probabilities, close to the extreme case, the unique invariant measure can be approximated by a smooth probability density.

1.3 Non-Deterministic Dynamical System

Given a hyperbolic IFS with probabilities $\{X; f_1, f_2, \ldots, f_N; p_1, p_2, \ldots, p_N\}$ the corresponding *non-deterministic dynamical system* is the iterative orbit of a single point in X, in which at each iteration a map f_i is selected with probability p_i.

In 1986, Elton [7] made the important observation that the frequency of visits to any given Borel subset of X is described by the unique invariant measure of $\{X; f_1, f_2, \ldots, f_N; p_1, p_2, \ldots, p_N\}$.

Using this observation, we are able to demonstrate computationally, by considering ordits of length 10^9, that the unique invariant measure gets closer to the formulated smooth probability density.

We then proceed to formulate the average Lyapunov exponent for the non-deterministic dynamical system and apply it to the problem of evaluating the storage capacity of forgetful neural networks using the smooth learning scheme.

2 The Approximate Probability Density

Consider the non-deterministic dynamical system with small ϵ corresponding to the hyperbolic iterated function system with probabilities $\left\{[x_-, x_+]; \phi_+, \phi_-; \frac{1}{2}, \frac{1}{2}\right\}$ where $\phi_\pm(x;\epsilon) = \phi(x \pm \epsilon)$, $\phi : \mathbb{R} \to \mathbb{R}$, ϕ is smooth, odd, monotonically increasing, strictly concave for $x > 0$, $\phi'(0) = 1$, $\epsilon > 0$ and $x_\pm$ is the fixed point of $\phi_\pm$. This prescription determines the function ϕ as having the form

$$\phi(x) = x - ax^r + O(x^{r+2}), \tag{1}$$

where $r \geq 3$ is an odd number and $a > 0$.

In general, the *order* notation $f(x) = O(g(x))$ means that for all $\delta > 0$, there exists a $C > 0$ such that $|x| < \delta$ implies $|f(x)| < C\,|g(x)|$.

Strictly speaking, the above IFS needs to be folded once to make it hyperbolic. In other words, we ought to consider

$$\left\{[x_-, x_+]; \phi_+ \circ \phi_+, \phi_+ \circ \phi_-, \phi_- \circ \phi_+, \phi_- \circ \phi_-; \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}$$

because there exists $c(\epsilon) < 1$ such that $\frac{d}{dx}\phi_\alpha \circ \phi_\beta(x) \leq c$ for all $\alpha, \beta \in \{+, -\}$, $x \in [x_-, x_+]$ and $\epsilon > 0$.

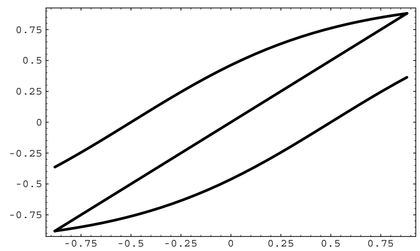

Figure 1: Plot of $y = \phi_+(x)$, $y = \phi_-(x)$ and $y = x$ from $x = x_-$ to $x = x_+$ where $\phi(x) = \tanh(x)$ and $\epsilon = \frac{1}{2}$

The unique invariant measure μ of the above IFS with probabilities is a fixed point of the *Frobenius-Peron equation,*

$$\begin{aligned} F : \mathbf{M}X &\rightarrow \mathbf{M}X \\ F(\nu) &= \frac{1}{2}\nu \circ \phi_+^{-1} + \frac{1}{2}\nu \circ \phi_-^{-1}. \end{aligned} \tag{2}$$

In other words,

$$\mu(B) = \frac{1}{2}\mu\left(\phi_+^{-1}(B)\right) + \frac{1}{2}\mu\left(\phi_-^{-1}(B)\right)$$

for all Borel sets $B \subseteq [x_-, x_+]$.

It is known that the shape of the invariant measure when represented by its value over certain narrow partitions of $[x_-, x_+]$ becomes gradually smoother [4] as $\epsilon \rightarrow 0$. However, it also becomes narrower. In fact, it tends to the Dirac delta function [5].

Let us change variables according to the prescription

$$y = \left(\frac{a^{\frac{1}{r+1}}}{\epsilon^{\frac{2}{r+1}}} \right) x \text{ and } h = a^{\frac{1}{r+1}} \epsilon^{\frac{r-1}{r+1}}. \tag{3}$$

This leads to a new IFS with probabilities $\{[y_-, y_+]; \psi_+, \psi_-; \frac{1}{2}, \frac{1}{2}\}$ where $\psi_\pm(y; h) = \psi(y \pm h)$, $y_\pm$ is the fixed point of $\psi_\pm$,

$$\begin{aligned} \psi : \mathbb{R} &\rightarrow \mathbb{R} \\ \psi(y; h) &= y - y^r h^2 + O(h^s) \end{aligned}$$

and $s = 2\left(\frac{r+1}{r-1}\right) > 2$.

It can be shown that for small ϵ, $x_{\pm} \approx \pm\sqrt[r]{\frac{\epsilon}{a}}$ and therefore that for small h, $y_{\pm} \approx \pm\sqrt[r]{\frac{1}{h}}$, which tends to $\pm\infty$ as $h \to 0$.

Another view of the invariant measure is the following. Imagine that we begin with an infinite number of starting points distributed according to a smooth probability density $p_0(x)$. The relation evolving the probability density forward in time is

$$p_{n+1}(y) = \int_{y_-}^{y_+} \left(\frac{1}{2}\delta(y - \psi_+(z)) + \frac{1}{2}\delta(y - \psi_-(z)) \right) p_n(z)\mathrm{d}z$$

where $\delta(y)$ is the Dirac delta function [5]. The unique invariant probability density $p(y)$ is the fixed point of the above iterative prescription.

Using the identity

$$\int_{-\infty}^{\infty} f(z)\delta(y - g(z))\mathrm{d}z = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}$$

we have

$$2p(y) = \frac{p(\psi_+^{-1}(y))}{\psi'_+(\psi_+^{-1}(y))} + \frac{p(\psi_-^{-1}(y))}{\psi'_-(\psi_-^{-1}(y))} \tag{4}$$

assuming that we can ignore that part of the integral where $y > y_+$ and $y < y_-$, which we certainly can do if $\phi(x)$ is bounded.

Furthermore, it can be shown that

$$\begin{aligned}
\psi_{\pm}(y;h) &= y \pm h - y^r h^2 + O(h^q) \\
\psi'_{\pm}(y;h) &= 1 - r y^{r-1} h^2 + O(h^q) \\
\psi_{\pm}^{-1}(y;h) &= y \mp h + y^r h^2 + O(h^s)
\end{aligned}$$

where $q = \min\{3, s\}$ and the differentiation and inversion are with respect to y.

Finally, by considering Taylor's expansion for the main parts on the right hand side of Eq. 4

$$\begin{aligned}
p(\psi_{\pm}^{-1}(y)) &= p(y) + (\mp h + y^r h^2)p'(y) + \frac{h^2}{2}p''(y) + O(h^s) \\
\psi'_{\pm}(\psi_{\pm}^{-1}(y)) &= 1 - r y^{r-1} h^2 + O(h^q)
\end{aligned}$$

and dividing through by h^2, we arrive at

$$p''(y) + 2y^r p'(y) + 2ry^{r-1}p(y) = O(h^{\lambda}) \tag{5}$$

where $\lambda = \frac{4}{r-1} > 0$.

Remarkably, this linear, second-order, differential equation has the exact solution

$$p(y) = K \exp\left(-\frac{2}{r+1} y^{r+1}\right) \tag{6}$$

as $h \to 0$ where K is an arbitrary constant. This solution goes to zero exponentially fast as $y \to \pm\infty$, perhaps justifying the earlier assumption about ignoring the far reaches of the integral range in the unbounded $\phi(x)$ case.

In order to satisfy the normalizing constraint for a probability density, we have to set

$$K = \frac{\left(\frac{r+1}{2}\right)^{\frac{r}{r+1}}}{\Gamma\left(\frac{1}{r+1}\right)} \tag{7}$$

where $\Gamma(x)$ is the gamma function [5], which can be defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t)\ \mathrm{d}t.$$

3 The Experimental Evidence

We backed up our careful mathematical analysis with computer simulations for a few cases of r and ϵ.

We considered $r = 3$ and $r = 5$. For each case, we started with $\epsilon = \frac{1}{2}$ and then halve it at each step until we reached $\epsilon = \frac{1}{32}$. For each step, we plotted the chaotic probability measure in black and the corresponding smooth probability density given by Eq. 6 in grey. The convergence of the two as ϵ decreases is striking.

The chaotic probability measure has been represented by partitioning the interval where the theoretical density is greater than one thousandth its value at zero into 500 equal sub-intervals and considering the frequency with which the orbit of the point starting at the origin visits each sub-interval over an excessive sequence of 10^9 iterations of the non-deterministic dynamical system.

Fig. 2 shows some of the plots for $\phi(x) = \tanh(x) = x - \frac{1}{3}x^3 + O(x^5)$ and Fig. 3 shows some of the plots for $\phi(x) = x - x^5$ with the horizontal axis representing x. Observe that the black lines are bounded exactly by x_- and x_+ as expected, while the grey lines are clearly not.

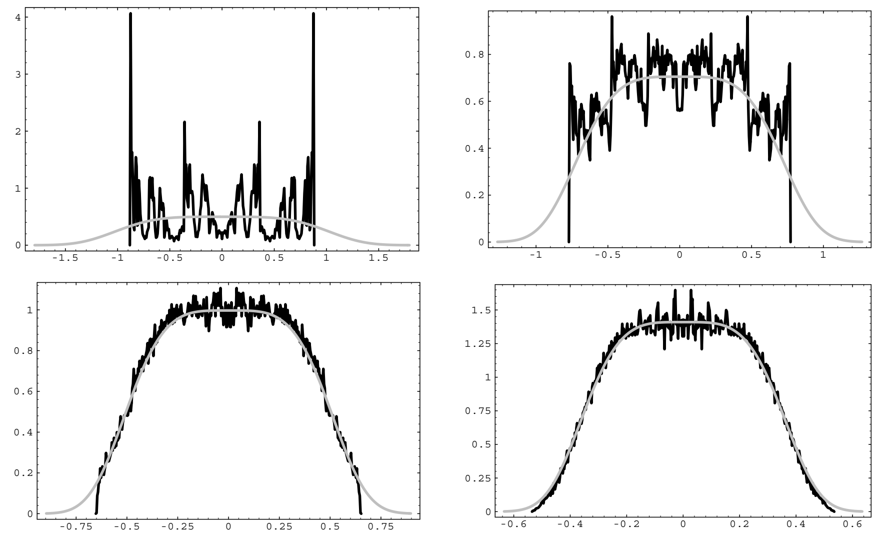

Figure 2: Plots of experimental measure (black) and theoretical density (grey) generated by IFS with $\phi(x) = \tanh(x)$ and $\epsilon = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ and $\frac{1}{16}$

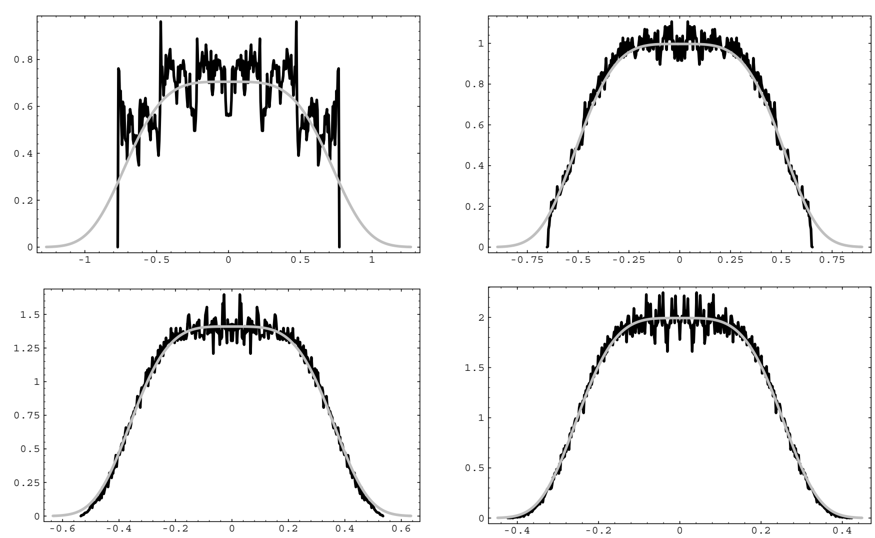

Figure 3: Plots of experimental measure (black) and theoretical density (grey) generated by IFS with $\phi(x) = x - x^5$ and $\epsilon = \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ and $\frac{1}{32}$

4 Analytical Expression for the Integral of any Smooth Function

Consider the integral

$$\int_{[x_-,x_+]} g(x)\ \mathrm{d}\nu(x) \tag{8}$$

where $g(x)$ is smooth and there exists n even such that the n^{th} derivative of $g(x)$ is its first non-zero derivative at the origin. There is no constructive technique for computing this integral [17] in general.

In the specific case, where ν is the unique invariant measure of a hyperbolic IFS with probabilities, a constructive technique, called *generalized Riemann integration*, by Edalat [6] can be used to compute the integral.

However, for the extreme case that we are considering in this paper the technique becomes impractical. In any case, it gives a numerical rather than an analytical answer.

Using the change of variable given in Eq. 3, we get

$$\begin{aligned}
\int_{[x_-,x_+]} g(x)\ \mathrm{d}\mu(x) &\approx \int_{-\infty}^{\infty} p(y) g\left(\left(\frac{\epsilon^{\frac{2}{r+1}}}{a^{\frac{1}{r+1}}}\right) y\right)\ \mathrm{d}y \\
&\approx \frac{g^{(n)}(0)}{n!} \frac{\Gamma\left(\frac{n+1}{r+1}\right)}{\Gamma\left(\frac{1}{r+1}\right)} \left(\frac{r+1}{2a}\epsilon^2\right)^{\frac{n}{r+1}}.
\end{aligned} \tag{9}$$

5 Analytical Expression for the Average Lyapunov Exponent

Consider two orbits, one starting at x and another starting at $x+\delta x$ where δx is small. The exponential rate at which these two orbits repel each other is known as the *Lyapunov exponent* γ_r at x. However, for the case in question, we have exponential attraction, therefore we would expect a negative Lyapunov exponent. In other words,

$$|\phi_{i_n} \circ \cdots \circ \phi_{i_2} \circ \phi_{i_1}(x+\delta x) - \phi_{i_n} \circ \cdots \circ \phi_{i_2} \circ \phi_{i_1}(x)| \approx \delta x \exp(\gamma_r n).$$

where $i_1, i_2, \ldots, i_n \in \{+,-\}$, n is large and δx is small.

It follows that

$$\gamma_r = \lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^{n} g(\phi_{i_m} \circ \cdots \circ \phi_{i_2} \circ \phi_{i_1}(x)) \tag{10}$$

where $g(x) = \ln \phi'_+(x)$ and $\phi(x)$ is odd.

The average Lyapunov exponent of a physical system quantifies its chaotic behavior. It has been found to be an extremely useful quantity to predict the macroscopic properties of many complex systems.

Elton's Ergodic Theorem [7] states:

> For a non-deterministic dynamical system corresponding to an IFS with probabilities $\{X; f_1, \ldots, f_N; p_1, \ldots, p_N\}$, where X is a compact metric space, the corresponding time average of a continuous function g for almost all initial points $x \in X$ and for almost all sequences $i_1, i_2, \ldots \in \{1, \ldots, N\}$ tends with probability one to its integral with respect to the unique invariant measure of the IFS
>
> $$\lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^{n} g(f_{i_m} \circ \cdots \circ f_{i_2} \circ f_{i_1}(x)) = \int g \, \mathrm{d}\nu \tag{11}$$
>
> provided that there exists $r < 1$ such that
>
> $$\prod_{i=1}^{N} d(f_i(x), f_i(y))^{p_i} \leq r d(x, y) \tag{12}$$
>
> for all $x, y \in X$ where d is the metric on X.

The condition for Elton's Ergodic Theorem is met in our case because $\phi_\pm$ are contracting maps and therefore the average Lyapunov exponent is given by

$$\gamma_r = \int g(x) \, \mathrm{d}\mu(x). \tag{13}$$

For small x, we have $g(x) \approx -arx^{r-1}$, and so, using Eq. 9, we have

$$\gamma_r \approx -ra^{\frac{2}{r+1}} \frac{\Gamma\left(\frac{r}{r+1}\right)}{\Gamma\left(\frac{1}{r+1}\right)} \left(\frac{r+1}{2}\epsilon^2\right)^{\frac{r-1}{r+1}}. \tag{14}$$

In particular, for $\phi(x) = \tanh(x)$, we have $\gamma_3 = -0.827901\epsilon$ and for $\phi(x) = x - x^5$, we have $\gamma_5 = -2.10909\epsilon^{\frac{4}{3}}$.

6 An Application to Forgetful Neural Networks

Neural networks [8] are the study of idealized systems containing very large numbers of connected neurons deliberately constructed to make use of organizational principles found in the human brain.

Forgetting in neural networks has a useful stabilizing role. Using the results above, we will derive an analytical expression for the optimal rate of forgetting. Optimal in the sense of maximizing its storage capacity for memorizing patterns.

Firstly, we need to consider the Hopfield model [9] with N fully connected neurons. Each neuron is a processing unit with one output $x_i \in \{-1, 1\}$.

The symmetric synaptic coupling parameter J_{ij} characterizes the connection from neuron j to neuron i.

Each neuron updates its output asynchronously according to

$$x_i \text{ becomes } \begin{cases} 1 & \text{if } \sum_{j=1}^{N} J_{ij} x_j > 0 \\ -1 & \text{otherwise} \end{cases} .$$

If we define the energy of the model as

$$H = -\frac{1}{2} \sum_{i,j=1}^{N} J_{ij} x_i x_j$$

it can be shown that it decreases monotonically to the local minimum.

We want to store binary patterns in the neural network. The obvious approach is to make the the required stored patterns correspond to local minima. This way when the neural network is set in motion close a local minimum, it should make it way down the energy mountain to this local minimum and hence act out a form of pattern recognition.

So, we require J_{ij} such that there are local minima corresponding to the required stored patterns. Given M patterns $\{\mathbf{X}^m \mid 1 \leq m \leq M\}$, the simplest storage prescription is

$$J_{ij} = \frac{1}{N} \sum_{m=1}^{M} X_i^m X_j^m .$$

Unfortunately, this prescription leads to *catastrophic forgetting* [1,2] for $M > 0.14N$.

Hopfield suggested alternative learning schemes in his original paper [9] that avoided this catastrophic forgetting. In these learning schemes, new patterns are learned at the expense of gradually forgetting previously stored patterns. These models are called *forgetful neural networks*. However, there is a price to pay for this more desirable behavior. Namely, the storage capacity is lower than that for the Hopfield model.

The forgetful storage prescription is given by the local iterative procedure

$$J_{ij}^m = \frac{1}{N} \phi(N J_{ij}^{m-1} + \epsilon X_i^m X_j^m) \tag{15}$$

where $J^0_{ij} = 0$ and $J_{ij} = J^M_{ij}$.

Let us consider the synaptic couplings. If we set $x_m = NJ^m_{ij}$ and $h_m = X^m_i X^m_j$ [4] then Eq. 15 reduces to

$$x_{m+1} = \phi(x_m + \epsilon h_m)$$

where $x_0 = 0$.

Assuming that the stored patterns are random, this means that h_m is a random variable, which is equal to 1 or -1 with equal probability. Therefore, in the limit as $M \to \infty$, this is a non-deterministic dynamical system corresponding to the IFS [10] with probabilities $\{[x_-, x_+]; \phi_+, \phi_-; \frac{1}{2}, \frac{1}{2}\}$ where $\phi_\pm(x) = \phi(x \pm \epsilon)$ and $x_\pm$ is the fixed point of $\phi_\pm$.

The nature of ϕ determines the properties of the neural network.

In 1986, Nadal et al. [13] explored various learning schemes for forgetful neural networks including the *marginalist* scheme and the *smooth* scheme.

The marginalist learning scheme specifies that the contribution of each pattern to the synaptic couplings decays exponentially with age.

$$\phi(x) = \Lambda\left(\frac{1}{N}\right) x \text{ where } \Lambda(z) = \epsilon \exp\left(-\frac{1}{2} z \eta^2\right). \tag{16}$$

In 1986, Mézard et al. [12] formulated and solved a general learning scheme that incorporated both the Hopfield model and the marginalist learning scheme in the thermodynamic limit as $N \to \infty$. It was shown that the maximum storage capacity is attained with $\eta_{\text{opt}} = 4.10812$ to 6 significant figures.

The smooth learning scheme specifies a general iterative procedure for the synaptic couplings using a restricted class of functions $\phi : \mathbb{R} \to \mathbb{R}$ such that ϕ is odd, monotonically increasing, strictly concave for $x > 0$ and $\phi'(0) = 1$. Notice that this prescription determines the function ϕ as having the form in Eq. 1.

The *embedding strength* of a stored pattern is a measure of how well represented it is in the synaptic couplings. Given M patterns $\{\mathbf{X}^m \mid 1 \leq m \leq M\}$, the embedding strength e_m of pattern $\mathbf{X}^m$ is defined by

$$e_m = \frac{1}{N} \sum_{i,j=1}^{N} J_{ij} X^m_i X^m_j. \tag{17}$$

We know that the embedding strengths of all stored patterns decay to zero as further patterns are subsequently stored by virtue of the properties of a forgetful neural network. Also, it would seem reasonable that this decay is exponential on average when large numbers of patterns have been previously stored because all patterns are homogeneous in character. This leads effectively to investigating the Lyapunov stability of the neural network, where the decay rate corresponds to the *Lyapunov exponent* [16,4].

Therefore, in the limiting case as $n \to \infty$, we have the following approximation up to scaling

$$e_m \sim \exp(\gamma_r n) \tag{18}$$

where $n = M + 1 - m$ and γ_r is the average Lyapunov exponent.

It can be shown [14] that for the marginalist learning scheme

$$e_m \approx \Lambda\left(\frac{n}{N}\right).$$

Comparing this with Eq. 18 up to scaling, we deduce that the optimal value of ϵ for a forgetful neural network with large N is

$$\epsilon_{\text{opt}} = \frac{\eta^2_{\text{opt}}}{2N\sqrt{6}} \frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} \approx \frac{10.19}{N} \tag{19}$$

for the canonical case where $\phi(x) = \tanh(x)$. Remember, it is not the actual values of the embedding strengths that count, but rather their values relative to each other.

Note that the optimal ϵ is only inversely proportional to N in the cases where $r = 3$. This is in contrast to that proposed in previous literature [16,4].

For the general case

$$\epsilon_{\text{opt}} \sim N^{-\frac{(r+1)}{2(r-1)}} \tag{20}$$

7 Conclusion

We have derived a smooth analytical expression Eq. 9 to approximate the integral of any smooth function over multifractal measures generated by iterated function systems with probabilities on the edge of chaos.

We have backed this up with experimental evidence and also applied it to the problem of calculating the storage capacity of forgetful neural networks.

There ought to be applications for this technique in statistical physics, and in particular to the Ising model.

References

1. D. J. Amit, H. Gutfreund, and H. Sompolinsky. Information storage in neural networks with low levels of activity. *Phys. Rev. A*, 35:2293–2303, 1987.

2. Daniel J. Amit. *Modelling Brain Function.* Cambridge University Press, 1992.
3. M. F. Barnsley and S. Demko. Iterated function systems and the global construction of fractals. *The proceedings of the Royal Society of London*, A399:243–275, 1985.
4. U. Behn, J. L. van Hemmen, R. Kühn, A. Lange, and V. A. Zagrebnov. Multifractality in forgetful memories. *Physica D*, 68:401–415, 1993.
5. R. Creighton Buck. *Advanced Calculus.* McGraw-Hill Book Company, 1965.
6. A. Edalat. Domain theory and integration. *Theoretical Computer Science*, 151:163–193, 1995.
7. John H. Elton. Ergodic theorem for iterated maps. *Ergod. Th. & Dynam. Sys.*, 7:481–488, 1987.
8. John Hertz, Anders Krogh, and Richard G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company, 1993.
9. J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 79:2554–2558, 1982.
10. J. E. Hutchinson. Fractal and self-similarity. *Indiana University Mathematics Journal*, 30:713–747, 1981.
11. B. Mandelbrot. *The Fractal Geometry of Nature.* W. H. Freeman and co., San Francisco, 1982.
12. M. Mézard, J. P. Nadal, and G. Toulouse. Solvable models of working memories. *J. Phys.*, 47:1457–1462, 1986.
13. J. P. Nadal, G. Toulouse, J. P. Changeux, and S. Dehaene. Networks of formal neurons and memory palimpsests. *Europhys. Lett.*, 1:535, 1986.
14. Peter J. Potts. The storage capacity of forgetful neural networks. Master's thesis, Department of Computing, Imperial College, Sept. 1995.
15. W. A. Sutherland. *Introduction to Metric and Topological Spaces.* Oxford University Press, 1993.
16. J. L. van Hemmen, G. Keller, and R. Kühn. Forgetful memories. *Europhys. Lett.*, 5:663–668, 1988.
17. Alan J. Weir. *Lebesgue Integration & Measure.* Cambridge University Press, 1994.