The Basins of Attraction of a New Hopfield Learning Rule

Amos J Storkey, Romain Valabregue

Neural Systems Group

Department of Electrical and Electronic Engineering

Imperial College

University of London, UK

The authors would like to thank P. de Wilde for helpful commments

Reprint requests to: Amos Storkey, now at Institute for Adaptive and Neural Computation, Division of Informatics, University of Edinburgh, 5 Forrest Hill, Edinburgh, EH1 2QL, UK. +44 131 651 1208.

amos@dai.ed.ac.uk

The Basins of Attraction of a New Hopfield Learning Rule

Abstract

The nature of the basins of attraction of a Hopfield network is as important as the capacity. Here a new learning rule is re-introduced. This learning rule has a higher capacity than that of the Hebb rule, and still keeps important functionality, such as incrementality and locality, which the pseudo-inverse lacks. However the basins of attraction of the fixed points of this learning rule have not yet been studied.

Three important characteristics of basins of attraction are considered: indirect and direct basins of attraction, distribution of sizes of basins of attraction and the shape of the basins of attraction.

The results for the new learning rule are compared with those of the Hebb rule. The size of direct and indirect basins of attractions are generally larger for the new rule than for the Hebb rule, the distribution of sizes is more even, and the shape of the basins more round.

KEYWORDS: Hopfield neural networks, attraction basins, learning rules.

1 Introduction

Attractor networks such as Hopfield networks[Hopfield, 1982] are used as auto-associative content addressable memories. The aim of such networks is to retrieve a previously learnt pattern from an example which is similar to, or a noisy version of, one of the previously presented patterns. To do this the network associates each element of a pattern with a binary neuron. These neurons are fully connected, and are updated asynchronously and in parallel. They are initialised with an input pattern, and the network activations converge to the closest learnt pattern.

In order to perform in the way described, the network must have a learning algorithm which sets the connection weights between all pairs of neurons so that it can perform this task.

These learning rules can have a number of characteristics. Firstly a rule can be local. If the update of a particular connection depends only on information available to the neurons on either side of the connection, then the rule is said to be local. Locality is important, because it provides a natural parallelism to the learning rule, which, when combined with the local update dynamics, make a Hopfield network a truly parallel machine.

Secondly a rule can be incremental. If the learning process can modify an old network configuration to memorise a new pattern, without needing to refer to any of the previously learnt patterns, then an algorithm is called incremental. Clearly incrementality makes the Hopfield network adaptive, and therefore more suitable for changing environments or real time situations.

Thirdly a rule can either perform an immediate update of the network configuration, or can be a limit process. The former makes for faster learning.

Fourthly a learning algorithm has a capacity. This is some measure of how many patterns can be stored in a network of a given size. The absolute capacity [McEliece et al., 1987] of a network is given by the number of patterns that can be stored with correct recall as an asymptotic function of the number of neurons.

Finally there is the concern of this paper: the sizes and shapes of the basins of attraction of the fixed points of the network. Patterns must not only be stored, but be able to be recalled. The recall process involves moving from the network start state into some fixed state. The basin of attraction of a fixed point is defined to be the set of states that are attracted to that fixed point. Different learning rules produce different basins of attraction. Clearly capacity and attraction basins are related. Here we consider what extra information about the performance of the new learning rule can be gained from the study of attraction basins.

The following are desirable features of the basins of attraction:

1. They should be large.
2. They should be rounded: states with similar Hamming distances from the fixed point should behave in the same fashion.
3. They should be evenly distributed.

Large attraction basins are desirable, because then more states are attracted to stored patterns rather than spurious fixed points. Round attraction basins make the network more predictable: whether or not a state is attracted to a stored pattern is reliably dependent on its Hamming distance away from that pattern. An even distribution of basin sizes also increases predictability and ensure there is no large bias against any particular pattern.

Attraction basins and capacity are not independent, and there have been methods to try to use the attraction basin size in a capacity-like definition. One such method can be found in [Schwenker et al., 1996]. There the concept of a completion capacity is introduced. This is an information theoretic concept, and it is used to calculate what is effectively the maximal information gain through the action of an associative memory for a given network size. Such a definition can be useful to represent the information processing power of a particular network. The analysis in [Schwenker et al., 1996] assumes both the independence of the components of the original patterns and the independence of the components after they have passed through the Hopfield update rule. This latter case is problematic, as the learning rule mixes the original independent components using the weight matrix. Hence the final components can be significantly correlated. A result of this problem is that the above approach, though generally useful, cannot be followed for the analysis undertaken here, such as that involved with looking at the shape and skew of attraction basins.

2 The Hopfield network

The Hopfield network is an attractor neural network governed by the difference equation

$$x_i(t+1) = \mathrm{sgn}(\sum_{j \neq i} w_{ij} x_j(t))$$

where $x_i(t)$ is the ± 1 state of neuron i, w_{ij} is the symmetric weight matrix and $x_i(n) = \pm 1$ is the nth update of the ith neuron where updates are performed asynchronously. The network is set up in an initial state, released, and the dynamics will take the network to the closest fixed point. The job of the learning rule of a Hopfield network is to find some weight matrix w_{ij} which stores the required patterns as fixed points of the network dynamics so that patterns can be recalled from noisy or incomplete initial inputs.

2.1 Different learning rules

Three learning rules are referred to in this paper. The Hebbian learning rule is local and incremental, but has a low absolute capacity of $m_c(n) = n/(2 \ln n)$ where n is the number of neurons [McEliece et al., 1987]. This capacity decreases significantly if patterns are correlated.

The pseudo-inverse rule [Kanter and Sompolinsky, 1987] has a higher capacity of $m_c(n) = n$, but does not have the functionality of the Hebb rule. It is not incremental or local, because it involves the calculation of an inverse. It has very small attraction basins for the higher network loadings.

2.1.1 The new learning rule

In order to overcome the problems of both of these learning methods we reintroduce the following new learning rule [Storkey, 1997].

Definition 1 (The new learning rule) *The weight matrix w_{ij} of an attractor neural network is said to follow the new learning rule if it obeys*

$$w_{ij}^0 = 0 \ \forall i,j \ \textit{and} \ w_{ij}^\nu = w_{ij}^{\nu-1} + \frac{1}{n}\xi_i^\nu \xi_j^\nu - \frac{1}{n}\xi_i^\nu h_{ji}^\nu - \frac{1}{n}h_{ij}^\nu \xi_j^\nu \qquad (1)$$

where $h_{ij}^\mu = \sum_{k=1,k\neq i,j}^n w_{ik}^{\mu-1}\xi_k^\mu$ is a form of local field at neuron i (the input to the neuron i) , and ξ^μ is the new pattern to be learnt ($\xi_i^\mu = \pm 1$)

This rule is local: w_{ij} depends only on information available at the two adjacent neurons (the values of ξ_i, ξ_j, h_i, h_j). It is clear from the recursive nature of (1) that it is also incremental. It has an absolute capacity of $m_c(n) = n/\sqrt{2\ln n}$ [Storkey, 1997] and performs better than the Hebb rule with correlated patterns [Storkey and Valabregue, 1997].

It should be noted that the weight matrix obtained is not independent of the presentation order. This important matter is discussed in another paper [Storkey, 1998]. There it is shown that this rule (or rather a minor variant of it) acts as a palimpsest or forgetful learning rule: it does not suffer from catastrophic forgetting when capacity is reached, but instead retains recent memories at the expense of old ones.

The dependence of the weight matrix on the presentation order will not affect the following analysis of macroscopic properties such as the size of attraction basins. This is because the patterns we will use will be statistically independent and identically distributed. Hence there is no structural order to the patterns, and they can perfectly well be interchanged without changing the relevant probability distributions.

2.2 The patterns

Here we consider training the network with the usual m patterns $\xi_i^\mu = \pm 1$ of length n $(i = 1, \ldots, n,\ \mu = 1, \ldots, m)$ with each bit chosen independently to be 1 with probability 1/2.

3 Direct and indirect basins of attraction

There are two fundamentally different types of basins of attraction. These are commonly called direct and indirect attraction basins. A state is within the direct basin of attraction of a fixed point if the dynamics of the network will always take the system state to the fixed point, and the network state will never move away from the fixed point throughout this process.

On the other hand a state not in the direct attraction basin, but which is eventually attracted to the fixed point (with a high probability) is said to be within the indirect attraction basin. We use the term complete attraction basin (or sometimes just attraction basin) to denote the union of the direct and indirect attraction basins.

It is one thing to know whether a state is within the attraction basin of a fixed point, and another thing to measure the size of the attraction basin. Here we represent attraction size by the hamming radius of the basin of attraction as is common in the literature [Forrest, 1988, Horner et al., 1989, McEliece et al., 1987, Viana and Coolen, 1993]. (Note that the hamming radius is equivalent to the overlap radius up to a linear transformation.) This in itself is not well defined, because attraction basins tend not to be Hamming hyper-spherical. One common measure is the minimum radius [Forrest, 1988, McEliece et al., 1987]:

$$r(p) = \inf\{d =< p, q >; p \in Basin(q)\} \tag{2}$$

where $< p, q >$ is the Hamming distance between states p and q in Q, and Q is the set of all stored patterns. Then to compare two learning rules, we can compare the average size of attraction basin over all the stored patterns

$$r = \frac{1}{|Q|} \sum_{p \in Q} r(p) \tag{3}$$

4 Direct basins of attraction

Direct attraction basins are important because the attraction rate is faster within the direct attraction basin than outside it. For synchronous dynamics, attraction occurs in one step. See section 7.3 for the reasons for this.

The definitions (2) and (3) were used to compare the size of direct attraction basins for different learning rules at different network loadings.

The direct attraction basins were calculated as follows:

- For each pattern, ξ^μ, find the state x with the shortest distance d from ξ^μ for which

  $$\sum_j w_{ij}\xi_i^\mu x_j$$

  is less than zero for some i. This implies that when the Hopfield update rule is applied to the neuron i of the network it returns $-\xi_i^\mu$, which is an incorrect update.

- The radius of the direct basin of attraction is therefore $d-1$.

The graph 1 shows the average direct basin sizes for three learning rules. Here a network size of 300 neurons was used. The graph for each learning rule terminates when absolute capacity is reached. The pseudo-inverse is included for comparison, even though it is neither a local nor incremental rule.

5 Indirect basins of attraction

Of course the direct basin of attraction is not of the greatest importance. The size of the complete attraction basin is that which most affects the ability of the network to perform as an associative memory. In this section we look at the following characteristics of complete attraction basins:

1. The distribution of sizes of attraction basins across different stored patterns.

2. The shapes of the attraction basins.

3. The effect that increasing the number of stored patterns has on the above factors.

Once again we need some way of measuring the sizes of the attraction basins. This is harder than it was with direct attraction basins. We chose a process similar to that used in [Forrest, 1988]. Our method is outlined below.

1. Choose some fixed point μ corresponding to a stored pattern.

2. Choose some initial normalised Hamming radius $r = r_0$.

3. Let the set A be all states of Hamming radius nr from the fixed point.

4. Sample 100 states from A.

5. Calculate how many of these states are attracted to the fixed point. Denote this number $t_\mu(r)$.

6. Increment the normalised Hamming radius r by a suitable amount and repeat from (2)

7. Repeat for each stored pattern, μ.

The benefit of this method is that $t_\mu(r)$ gives an estimate for the percentage of states distance nr from the fixed point μ which are attracted to the fixed point. We arbitrarily define the radius R of attraction to be the largest value of r such that $t_\mu(s) \geq 90$ for all $s \leq r$. In other words most states within a Hamming sphere radius R are attracted to the fixed point. Note that $t_\mu(r)$ can also give a measure of the skew of the basin of attraction. The skew of an attraction basin is some measure of how much it differs in shape from a hypersphere. The more elongated or contorted, the greater the skew.

If $t_\mu(r)$ decreases rapidly as r increases from R then the basin of attraction is approximately hyperspherical: at one radius almost all states are attracted, and at a slightly higher radius very few are attracted. If, however, the gradient of $t_\mu(r)$ is shallow, then the basin of attraction is skew: Some states are attracted from further away than others.

6 Results

The graphs of figure 2 give the values of $t_\mu(r)$ for each pattern μ, for a network size of 150 neurons, and a number of different network loadings.

From such results we can compare the size of the complete attraction basins for the Hebb rule and the new learning rule. Figure 4 gives the distributions of attraction basin sizes for each of the rules for different loadings.

Lastly we look at how the skew of each attraction basin varies with the size of the attraction basin. Here, skew is defined as $t_\mu^{-1}(90) - t_\mu^{-1}(40)$. This slightly arbitrary, but simple definition suffices to approximate the gradient of $t(r)$. The larger the skew, the greater the variety of distances from which states are attracted to the stored pattern. This implies that the attraction basin is less rounded. Figure 3 gives a comparison of the skew for the Hebb rule and the new rule. Note that different loadings (m/n) need to be used for the different rules. This is because the network needs to be near capacity to get any variation in basin size, and the capacity of the new rule is greater than that of the Hebb.

7 Analysis

7.1 Basin Sizes

It is clear from figures 1 and 4 that the attraction basins are generally larger for the new rule than for the Hebb rule. The difference is there, but not highly significant at very low loadings. As the number of stored patterns increases this difference becomes more noticeable. Near the capacity of the new rule, the Hebb rule basin size for most stored patterns has collapsed. On the other hand the new rule still has large basin sizes for the majority of patterns.

7.2 Basin distribution and network predictability

Not only do the average radii of the attraction basins differ, but the spread of the distribution of radii is different. For the Hebb rule the distribution of attraction basin radii is peaked at low network loadings. (figure 4) As the Hebb rule capacity is approached, one of the attraction basins suddenly begins to loose size. The rest of the attraction basins remain in a peaked distribution. One by one, as the network loading is increased, other attraction basins 'leave the pack' and quickly loose size. At capacity the first attraction basin disappears. Long before the network reaches the capacity loading for the new rule, the attraction basins of the stored patterns have all disappeared.

The behaviour of the new rule is different. It also has a tight distribution of basin sizes for low network loadings. However, as the loading increases, attraction basins do not loose size quickly. Instead the distribution of basin sizes slowly spreads out. Capacity is reached when the lower tail of the distribution of attraction basin sizes touches zero, and one of the stored patterns is no longer an attractor. At capacity, most stored patterns still have substantial attraction basin sizes.

From this we ascertain that the behaviour of the new rule is more predictable than the Hebb rule: the distribution of basin sizes does not change dramatically for small changes in network loadings. Furthermore, the distribution of attraction basin sizes is more even: The distribution is a bell-shaped distribution with a standard deviation that increases slowly as the network loading increases, not a bimodal distribution like that of the Hebb rule near capacity.

7.3 Direct attraction basins

Not only are the complete attraction basins larger for the new rule than the Hebb rule, but the direct attraction basins are larger, and decay slower as m increases. Network performance is better within the direct attraction basins, because attraction rates are faster. Network states are attracted to the memory in one synchronous time step. In asynchronous terms this means that the attraction rate is exponential because at each step the probability of moving 1 hamming unit towards the fixed point is proportional to the distance from the current state to the fixed point, and the probability of moving away is zero. Outside the direct attraction basin the attraction rate is generally sub exponential, because the probability of moving away from the fixed point is non-zero.

We note that the direct attraction basins are not significantly different in size from those of the pseudo-inverse.

7.4 Shape of attraction basins

Lastly we consider the shape of the attraction basins. Attraction basins for the Hebb rule are not round but have a convoluted shape [Amari and Maginu, 1988]. This can create problems with misclassification because initial states which are close to a particular stored pattern are actually attracted to another pattern which is further away. More likely, states are attracted to spurious memories, and convoluted basins mean that some states close to the stored pattern are attracted to spurious states, whereas other states initially further away are attracted to the pattern. The results obtained are therefore highly unpredictable.

Note that in general, for a content addressable memory (CAM), hyperspherical attraction basins would not be expected. For example if the three patterns A,B and C are such that A and B are close, but A and B are far from C, then the basin of attraction about A would be smaller in the direction of B than in the direction of C.

However, the attraction basins of the Hebb rule attractors are not skew in the above way. This can be shown as follows:

Consider m random patterns $\xi_i^\mu = \pm 1$ defined as before. Let the distance between any two patterns μ and ν, $d(\mu,\nu)$ be the hamming distance.

The distribution of the hamming distance between any two patterns has mean $n/2$ and variance $n/4$. We want to find the probability that any two patterns are less than $n/2 - \alpha n$ apart or greater than $n/2 + \alpha n$ apart. We need both of these inequalities to account for the fact that the Hopfield model effectively stores both a pattern and its negative.

Proposition 1 *The probability that some pair of patterns, say ξ^a and ξ^b, can* *be found with* $|d(\xi^a, \xi^b) - n/2| > \alpha n$ *is less than*

$$\frac{(1/2-\alpha)^{(\alpha-1/2)n} m^2}{2^{n-1}(1/2+\alpha)^{(\alpha+1/2)n}}$$

Proof Let A_μ^ν be the event that $|d(\mu,\nu) - n/2| > \alpha n$. Then the event

$$A_\mu = \cup_{\nu \neq \mu} A_\mu^\nu$$

has probability

$$P(A_\mu) < \sum_{\nu \neq \mu} P(A_\mu^\nu) < mP(A_1^2)$$

using the fact that the distance between any two patterns is identically distributed. Also the event $A = \cup_\mu A_\mu$ has probability $P(A) < \sum_\mu P(A_\mu)$ and so $P(A) < m^2 P$ where $P = P(A_1^2)$

By symmetry $P(A_1^{2+}) \stackrel{\text{def}}{=} P[d(1,2) > n/2 + \alpha n] = P[d(1,2) < n/2 - \alpha n]$, and these two events are mutually exclusive. Hence $P = P(A_1^2) = 2P(A_1^{2+})$.

Now

$$\begin{aligned} P = 2P[d(1,2) > n/2 + \alpha n] &< \frac{2\mathbf{E}|\exp[td(1,2)]|}{\exp(\alpha nt + nt/2)} \text{ for all } t \\ &= \frac{2(1+e^t)^n}{2^n e^{\alpha nt} e^{nt/2}} \end{aligned}$$

by a Chebyshev inequality [Grimmett and Stirzaker, 1982, p 186]. We have used the fact that the moment generating function of a binomial$(n, 1/2)$ distribution is $(1+e^t)^n/2^n$. By equating the derivative of P with respect to t with zero, we find that the P is minimised by choosing $t = \ln(1/2+\alpha) - \ln(1/2-\alpha)$. Substituting this in and using $P(A) < m^2 P$ we get the required result. □

In our case, for the Hebb rule, we have $n = 250$, $m = 19$. If we take $\alpha = 0.2$, then $P(A) < 9 \times 10^{-7}$. In addition, this probability tends to zero as $n \to \infty$ with $m = o(n)$.

The above shows the probability that the distance between any two patterns is significantly different (more than $0.2n$) from $n/2$ is negligible. This means that the radius of attraction should not decrease below $0.3n$ in any direction, if the variation in the radius is due to the above effect. Furthermore the skew of patterns should not be more than 0.1: With the above effect, the midpoint between two patterns defines the boundary of the basins of attraction to each point (assuming no closer pattern). The distance from a pattern to the midpoint is half the distance between the two points.

The noticed shape of the attraction basins for the Hebb rule cannot be attributed to the above effect. The skew is often larger than 0.1 and the radii of attraction basins is often well below 0.3 in *all* directions (see figures 3 and 4). We find that the attraction basins for the Hebb rule become more skew as they become smaller. For the new rule, the attraction basins tend to remain quite rounded, whatever their size.

7.5 Other details

The use of a fixed sample size of 100 might seem small for this application. However because independent sampling is made every step in radius (of 2 hamming units), the real sample size is actually much larger: results for close radius sizes are expected to be similar. The difference between 5 independent runs of this technique was found to be insignificant.

The authors experimented with different network sizes. The results were similar, although the differences between the Hebb and new learning rules became more marked as the network sizes increased.

8 Conclusions

It is already known that the new learning rule has a higher capacity than the Hebb rule, and does not suffer significant capacity loss when patterns with medium correlation are stored.

In this paper we have established some initial results about the size and shapes of the attraction basins. Empirical methods (somewhat similar to those used in [Forrest, 1988]) have been used to gain knowledge of the sizes and shapes of the basins for independent unbiased storage patterns. This paper has shown that the new rule has larger, more evenly distributed and more rounded attraction basins than those of the Hebb rule.

This new rule provides a different approach to the problem of increasing capacity and the sizes of attraction basins. Other approaches include the introduction of dynamic or noisy external fields [Wang, 1994, Yau and Wallace, 1991, Yau and Wallace, 1992], self feedback [Ho and DeWilde, 1995] or the minimum overlap method of [Krauth and Mezard, 1987, Chang and Wu, 1993]. Here, however, we do not only focus on the sizes of the attraction basins, but the shape and distribution as well. The warped attraction basins created by the Hebb rule, and noted in [Amari and Maginu, 1988] are irradicated by this method.

The algorithm we propose achieves its results while maintaining important functionality: unlike the pseudo-inverse rule or minimum overlap rule, it is incremental and local, and it is relatively fast because it does not involve the calculation of an inverse of an m by m matrix.

It should be noted that we have not mentioned the external field issue in this paper, but just examined the learning rule as it stands. There is no reason that the use of external fields cannot also be applied to this rule to increase attraction basin sizes.

References

[Amari and Maginu, 1988] Amari, S. I. and Maginu, K. (1988). Statistical neurodynamics of associative memory. *Neural Networks*, 1:63–73.

[Chang and Wu, 1993] Chang, J. Y. and Wu, C. C. (1993). Desing of hopfield-type associative memory with maximal basin of attraction. *Electronics Letters*, 29(24):2128–2130.

[Forrest, 1988] Forrest, B. M. (1988). Content-addressability and learning in neural networks. *Journal of Physics A: Mathematical and General*, 21:245–255.

[Grimmett and Stirzaker, 1982] Grimmett, G. R. and Stirzaker, D. R. (1982). *Probability and Random Processes*. Oxford University Press.

[Ho and DeWilde, 1995] Ho, A. M. C. L. and DeWilde, P. (1995). General transient length upper bound for recurrent neural networks. *Lecture Notes in Computer Science*, 930:202–208.

[Hopfield, 1982] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America: Biological Sciences*, 79(8):2554–2558.

[Horner et al., 1989] Horner, H., Bormann, D., Frick, M., Kinzelbach, H., and Schmidt, A. (1989). Transients and basins of attraction in neural network models. *Zeitschrift fur Physik B*, 76:381–398.

[Kanter and Sompolinsky, 1987] Kanter, I. and Sompolinsky, H. (1987). Associative recall of memory without errors. *Physical Review A: General Physics*, 35(1):380–392.

[Krauth and Mezard, 1987] Krauth, W. and Mezard, M. (1987). Learning algorithms with optimal stability in neural networks. *Journal of Physics A - Mathematical General*, 20(11):L745 –L752.

[McEliece et al., 1987] McEliece, R. J., Posner, E. C., Rodemich, E. R., and Venkatesh, S. S. (1987). The capacity of the Hopfield associative memory. *IEEE Transactions on Information Theory*, 33(4):461–482.

[Schwenker et al., 1996] Schwenker, F., Sommer, F. T., and Palm, G. (1996). Iterative retrieval of sparsely coded associative memory patterns. *Neural Networks*, 9:445–455.

[Storkey, 1997] Storkey, A. J. (1997). Increasing the capacity of the Hopfield network without sacrificing functionality. In Gerstner, W., Germond, A., Hastler, M., and Nicoud, J., editors, *ICANN97: Lecture Notes in Computer Science 1327*, pages 451–456. Springer-Verlag.

[Storkey, 1998] Storkey, A. J. (1998). Palimpsest memories: a new high-capacity forgetful learning rule for Hopfield networks. In preparation.

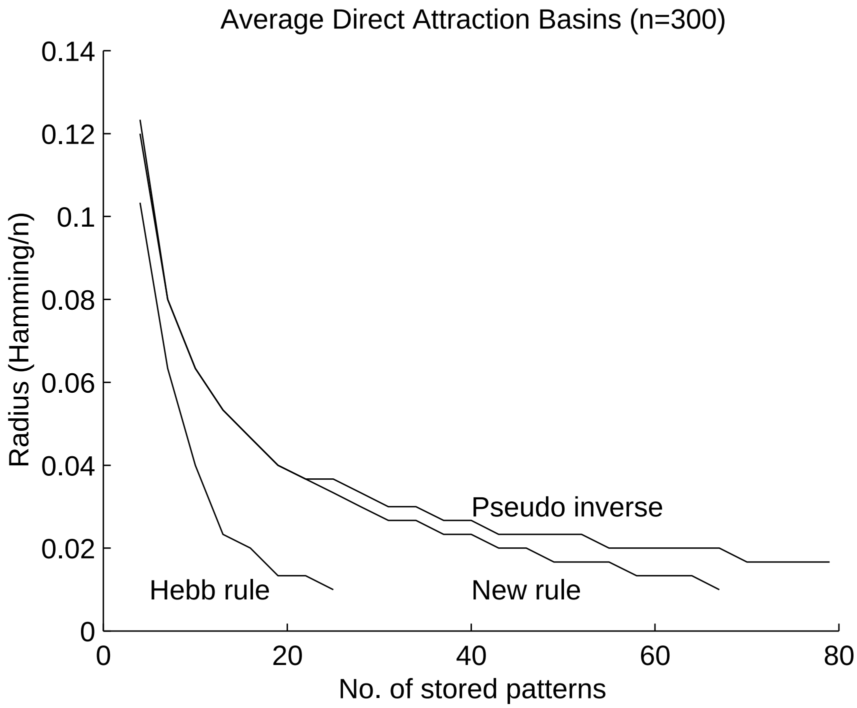
[Storkey and Valabregue, 1997] Storkey, A. J. and Valabregue, R. (1997). A new Hopfield learning rule with high capacity storage of correlated patterns. *Electronics Letters*, 33(21):1803–1804.

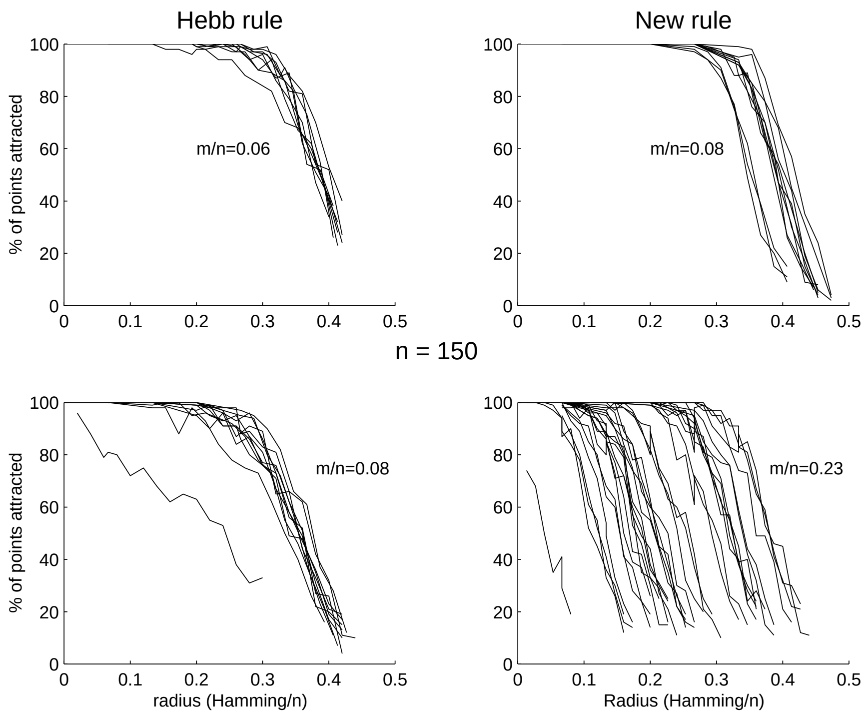
[Viana and Coolen, 1993] Viana, L. and Coolen, A. (1993). Attraction domains in neural networks. *Journal de Physique 1*, 3:777–786.

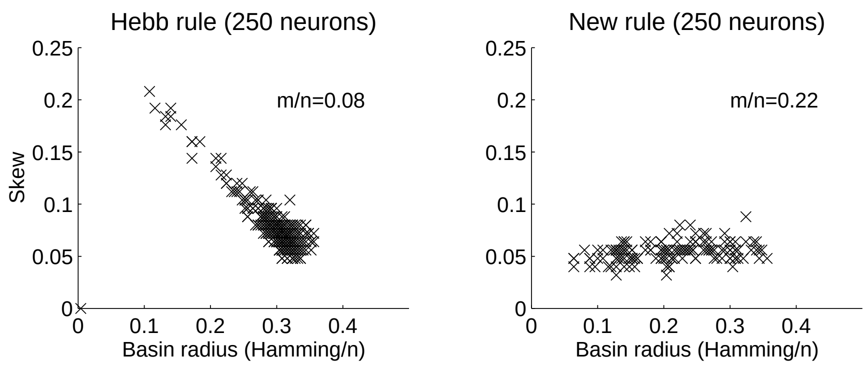
[Wang, 1994] Wang, T. (1994). Improving recall in associative memories by dynamic threshold. *neural networks*, 7(9):1379–1385.

[Yau and Wallace, 1991] Yau, H. W. and Wallace, D. J. (1991). Enlarging the attractor basins of neural networks with noisy external fields. *Journal of Physics A - Mathematicaland General*, 24(23):5639–5650.

[Yau and Wallace, 1992] Yau, H. W. and Wallace, D. J. (1992). Basins of attraction in a neural network model trained with external fields. *Physica A*, 185(1-4):471–480.


Average Direct Attraction Basins (n=300)
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Radius (Hamming/n)
0
20
40
60
80
No. of stored patterns
Pseudo inverse
Hebb rule
New rule


Hebb rule
New rule
% of points attracted
m/n=0.06
m/n=0.08
n = 150
m/n=0.08
m/n=0.23
radius (Hamming/n)
Radius (Hamming/n)


Hebb rule (250 neurons)
New rule (250 neurons)
m/n=0.08
m/n=0.22
Skew
Basin radius (Hamming/n)
Basin radius (Hamming/n)
0.25
0.2
0.15
0.1
0.05
0
0.1
0.2
0.3
0.4

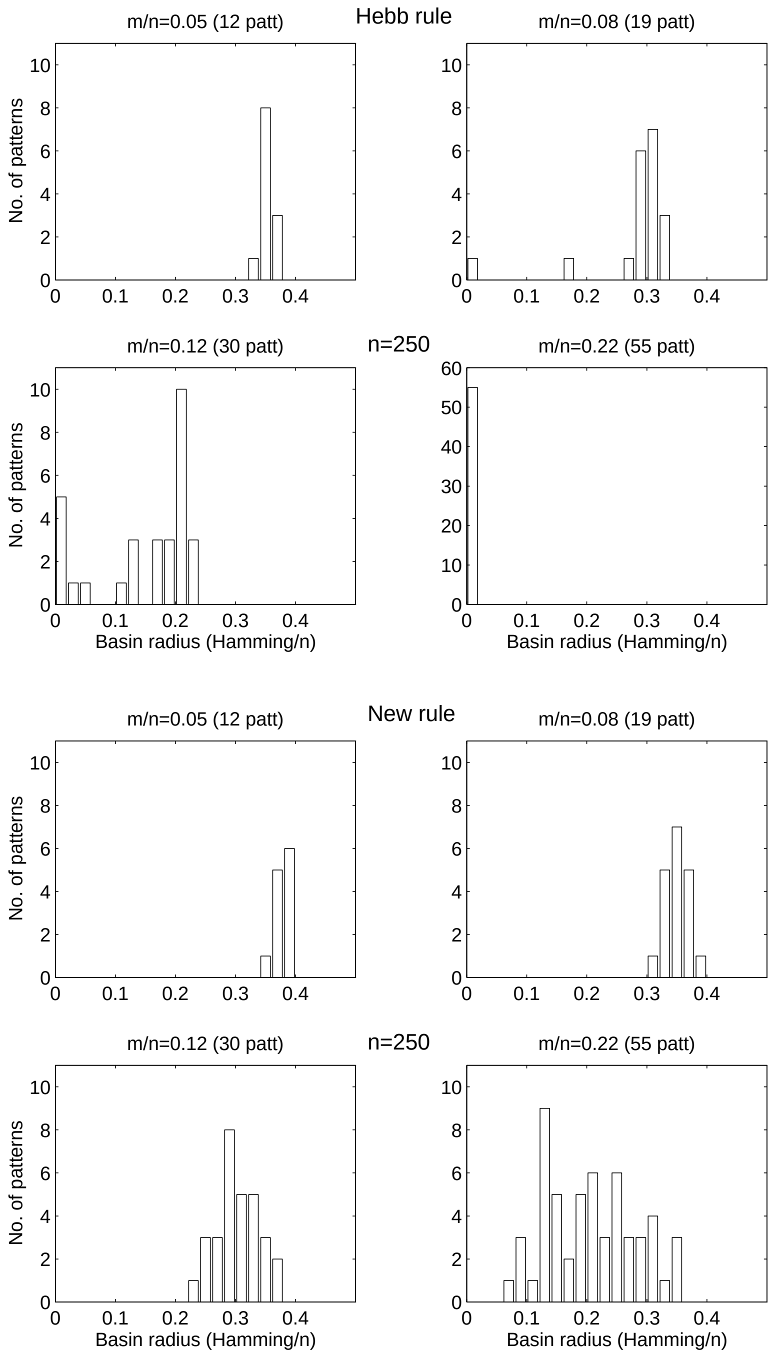

Hebb rule
m/n=0.05 (12 patt)
m/n=0.08 (19 patt)
n=250
m/n=0.12 (30 patt)
m/n=0.22 (55 patt)
New rule
m/n=0.05 (12 patt)
m/n=0.08 (19 patt)
n=250
m/n=0.12 (30 patt)
m/n=0.22 (55 patt)
No. of patterns
Basin radius (Hamming/n)
0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
20
30
40
50
60

Figure 1: Comparing the sizes of direct attraction basins

Figure 2: The proportion of states at radius r which are attracted $(t(r))$ for m stored patterns, and $n = 150$ neurons

Figure 3: Comparing the skew of attraction basins

Figure 4: Graph shows the distribution of attraction basin sizes across different stored patterns