The number of direct attractors in discrete state neural networks
R. Raju Viswanathan

To cite this version:

HAL Id: jpa-00246940
https://hal.archives-ouvertes.fr/jpa-00246940
Submitted on 1 Jan 1994

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The number of direct attractors in discrete state neural networks

R. Raju Viswanathan

Centre for Artificial Intelligence and Robotics, Raj Bhavan Circle, Bangalore 560001, India

(Received 15 September 1993, received in final form 28 January 1994, accepted 2 February 1994)

Abstract. — The maximum possible number of isotropic direct attractors of given Hamming radius in discrete state neural networks is obtained by means of evaluation of the fractional volume in the space of interactions. This number is proportional to the total number of neurons in the network.

An important problem in the study of any dynamical system is the classification of its fixed point attractors. If the system has such attractors, their distribution in the space of couplings and the approach to the corresponding equilibria are both of considerable interest. A knowledge of the distribution of various kinds of attractors in the space of couplings would tell us about what kind of distribution of attractors one could expect for a generic choice of couplings while information about the approach to the corresponding equilibria would tell us about the behaviour of neighbouring states. In general this problem of the 'phase structure' of attractors is difficult to study. In this paper, we shall study a particular class of dynamical systems, namely bi-state fully connected neural networks evolving in discrete time, and provide an answer to the following question regarding the distribution of a specific type of attractor—what is the maximum number of direct attractors, of given radius of attraction in the space of states, that one can expect to find? We shall answer this question in the limit of large $N$ (number of neurons), by evaluating the entropy in the space of interactions [1]. The issue of training with noise and the basin of attraction for sparsely connected networks (with the connectivity per neuron $C \ll \log N$) was studied in reference [2] by evaluating the fixed points of the evolution equation of the overlaps.

To be more specific, our system consists of $N$ neurons, each of which can exist in one of two possible states, $s_i = \pm 1$, for $i = 1, \ldots, N$. The phase space of possible states of the system then consists of $2^N$ points. The Hamming distance between two states A and B is the number of neurons whose states in A and B are different:

$$d(A, B) = \frac{1}{4} \sum_i (s_{iA} - s_{iB})^2.$$
As is usual, we shall use the following updating rule for the state of the \(i\)-th neuron at time \(t + 1\) [3]:

\[
s_i(t + 1) = \text{sign}\left(\sum_{j \neq i} w_{ij} s_j(t)\right)
\]

where \(w_{ij}\) (the synaptic weight) is the strength of the coupling from the \(j\)-th neuron to the \(i\)-th neuron. The only constraint we shall place on the couplings is that they be bounded [1,4]:

\[
\sum_{j \neq i} w_{ij}^2 = N.
\]

A direct attractor of radius \(R\) is defined as a fixed point \(\{s\}\) to which any other state \(\{s'\}\) at Hamming distance \(R\) from \(\{s\}\) converges in a single update:

\[
\text{sign}\left(\sum_{j \neq i} w_{ij} s'_j\right) = s_i,
\]

for all \(i\).

As was shown in reference [5], it is easy to see that if all states at a distance \(R\) from a given fixed point are attracted to the fixed point in a single update, then so are all states at distance \(R - 1\) from the fixed point. One way to see this is as follows. Suppose the \(i\)-th bit of the given fixed point, without loss of generality, is +1. Then for any such state \(s'\) at distance \(R\) from the fixed point, \(W_i s' > 0\), where \(W_i\) is the \(i\)-th row of the matrix of the weights. Now suppose a state \(s\) at distance \(R - 1\) from the fixed point differs from some state \(s'\) at distance \(R\) from the fixed point at one bit, say the \(n\)-th bit (here \(s_n = \sigma_n\), where \(\sigma\) is the fixed point; \(s_n\) is one of the \(N - R + 1\) bits at which \(s\) and \(\sigma\) agree). Then we have

\[
W_i s = W_i s' - 2w_{in} s'_n > -2w_{in} \sigma_n = 2w_{in} \sigma_n.
\]  

Now consider another state \(s''\) at a distance \(R\) from \(\sigma\) and distance 1 from \(s\). The condition \(W_i (s'' + \sigma) > 0\) tells us that \(\sum w_{ik} \sigma_k > 0\), where the sum is over those bits \(\sigma_k\) at which \(s''\) and \(\sigma\) (and \(s\) as well) agree; there are \(N - R\) such bits. Therefore at least one of the terms in this sum must be positive. We can always choose \(w_{in} \sigma_n\) to correspond to one of these terms which are positive (i.e., we can choose \(s'\) appropriately by simply flipping the \(n\)-th bit of \(s\)). Then from equation (2) above, we have \(W_i s > 0\), indicating that \(s\) is attracted to \(\sigma\) in a single update. Since \(s\) was arbitrary, it follows that all states at distance \(R - 1\) from \(\sigma\) are directly attracted to \(\sigma\), if all states at distance \(R\) from \(\sigma\) are directly attracted to \(\sigma\). In other words, isotropic direct attractors are surrounded by 'balls' about the fixed point, such that any state in the ball is attracted to the fixed point in a single update.

An immediate corollary to this result is the following: if \(p_n\) denotes the maximum possible number of direct attractors of radius \(n\), then \(p_R \leq p_{R-1}\). The maximum possible number of direct attractors of given radius of attraction is a non-increasing function of the radius of direct attraction.

We shall now proceed to calculate the partition function in the space of interactions [1,4]. To ensure that all states at a Hamming distance less than or equal to \(R\) are attracted to a given fixed point in a single iteration, it suffices to ensure that all points on the Hamming 'sphere', \(H_R\), of states at distance \(R\) from the fixed point are attracted to the fixed point in one update.
In particular, suppose \( \mathcal{W} \) is a region in weight space that stores \( p \) given patterns as direct attractors of radius \( R \). Let \( s'^\mu \) denote a state on the Hamming sphere of radius \( R \) about the attractor \( s^\mu \), for which the quantity

\[
K = \frac{1}{\sqrt{N}} \sum_{i,j} w_{ij} t'^\mu
\]

is least, among all indices \( i \), over all the states \( t'^\mu \) on \( H_R \), and among all points in \( \mathcal{W} \). Then clearly, if \( s'^\mu \) is attracted to \( s^\mu \) in a single update (which means \( K \) is bigger than some threshold value), so are all other states on \( H_R \), and \( s^\mu \) is then a direct attractor of radius \( R \).

Accordingly, the quantity we seek to minimise is the Hamiltonian

\[
H[w] = \sum_{i=1}^{N} \sum_{\mu=1}^{p} \left[ \theta (\kappa - \gamma_i^\mu [w]) + \theta (\kappa_1 - \gamma_{i1}^\mu [w]) \right]
\]

where

\[
\gamma_i^\mu = s_i^\mu \sum_{j \neq i} \frac{w_{ij} s_j'^\mu}{\sqrt{N}}
\]

and

\[
\gamma_{i1}^\mu = s_i^\mu \sum_{j \neq i} \frac{w_{ij} s_j'^\mu}{\sqrt{N}}.
\]

Here the \( s_i^\mu (\mu = 1, \ldots, p) \) are the \( p \) patterns to be stored as fixed point attractors, and \( s'^\mu \) is a state at Hamming distance \( R \) from \( s^\mu \) with the least value of \( K \) (see previous paragraph). The second \( \theta \)-function in \( H \) ensures that \( s^\mu \) is a fixed point, and the first ensures that \( s'^\mu \) is attracted to \( s^\mu \) in a single iteration. The entropy in interaction space is given by a quenched average over all the states \( s'^\mu \) corresponding to each pattern \( s^\mu \), followed by an average over the random patterns \( s \) themselves. Introducing a 'temperature' parameter \( \beta \), the partition function corresponding to this Hamiltonian is

\[
Z = \int \left( \prod_{j \neq i} dw_{ij} \rho_i[w] \right) e^{-\beta H[w]}
\]

where \( \rho_i[w] \) is the density of states in interaction space, which enforces the spherical constraint on the \( w \)'s mentioned earlier:

\[
\rho_i[w] = \delta \left( \sum_{j \neq i} w_{ij}^2 - N \right).
\]

As usual, in the \( \beta \rightarrow \infty \) limit, \( Z \) is simply the fractional volume of zero energy states, \( \Omega(0) \), so that in this limit the volumetric entropy of the zero energy state is

\[
S = \ln \Omega(0)
= \ln \int \left( \prod_{i} dw_{ij} \right) \prod_{\mu=1}^{N} \prod_{\mu=1}^{p} \theta (\gamma_i^\mu [w] - \kappa) \theta (\gamma_{i1}^\mu [w] - \kappa_1) \rho_i[w].
\]

The quantity we want is the entropy averaged over all the \( s'^\mu \) corresponding to the \( p \) stored patterns, and over the random patterns \( s^\mu \) themselves.
It is reasonable to assume that for random patterns \( s^\mu \) which are not correlated with each other, the states \( s'^\mu \) are uncorrelated with each other as well and have an overlap only with the corresponding fixed points \( s^\mu \). The expression for the averaged entropy factorises over the sites \( i \):

\[
\ll S \rr = \sum_i \ll \ln \Omega_i \rr
\]

with

\[
\Omega_i = \int \left( \prod_{\mu=1}^{P} d\gamma_{ij} \right) \theta(\kappa - \gamma_{ij}^\mu) \theta(\kappa_1 - \gamma_{1i}^\mu[w]) \rho_i[w].
\]

The replica trick can be used to evaluate \( \ll \ln \Omega_i \rr \). Replacing the \( \theta \)-functions by appropriate integrals over \( \delta \)-functions, and using integral representations for the delta functions, we note that all averages are to be carried out after expanding exponentials in the large-\( N \) limit and using \( \ll s_j^\mu s_k^\nu \rr = \ll s_j^\mu s_k^\nu \rr = \ll s_j^\mu s_k^\nu \rr = 0 \) for \( j \neq k \) and/or \( \mu \neq \nu \) (since we have assumed random patterns), and \( \ll s_j^\mu s_j^\nu \rr = (1 - 2f) \), where \( f = R/N \) is the fractional Hamming distance between states \( s \) and \( s' \). This last expression for the correlation between \( s^\mu \) and \( s'^\mu \) follows from the fact that since \( s'^\mu \) is at Hamming distance \( R \) from \( s^\mu \), they differ in \( R = fN \) bits; the probability for \( s_j^\mu \) to agree with \( s_j'^\mu \) is therefore \( (1 - f) \), and the probability for them to differ is \( f \).

Then one gets the following expression for the averaged entropy in the saddle point approximation:

\[
\ll S \rr = \lim_{n \to 0} \frac{\exp(N^2G) - 1}{n}
\]

where

\[
G = \alpha G_1(q_{ab}) + G_2(E_a, F_{ab}) + i \sum_{a < b} q_{ab} F_{ab}.
\]

(4)

Here \( a \) and \( b \) are replica indices, and \( F_{ab} \) and \( E_a \) are order parameters which serve to impose the constraint

\[
q_{ab} = \frac{1}{N} \sum_j w_{ij}^a w_{ij}^b
\]

on the order parameter \( q_{ab} \) and the spherical constraint on the weights respectively. \( \alpha = p/N \) is the storage capacity. Assuming the replica symmetric ansatz \( q_{ab} = q, F_{ab} = iF, E_a = iE, \) the saddle point conditions \( \frac{\partial G}{\partial E} = \frac{\partial G}{\partial F} = 0 \) yield

\[
E = \frac{1}{2} \frac{1 - 2q}{(1 - q)^2}
\]

and

\[
F = \frac{q}{(1 - q)^2}
\]

as usual. \( G_1 \) reduces, after changing variables appropriately and evaluating Gaussian integrals, to

\[
G_1 = \int_{-\infty}^{\infty} Dz \int_{-\infty}^{\infty} Dt \ln \left( \int_{t}^{\infty} DwH(y_0) \right),
\]

(5)

where

\[
y_0 = \frac{\kappa_1 + \sqrt{q(rz + st)}}{s\sqrt{1 - q}} - \frac{rw}{s},
\]
Fig. 1. — The maximum storage capacity as a function of the radius of attraction for \( \kappa_1 = 0, 1 \) and 2; the upper curves are for the smaller values of \( \kappa_1 \).

\[
\tau = \frac{\kappa + \sqrt{q z}}{\sqrt{1 - q}}, \quad \text{and } H(y) \text{ is the complementary error function } H(y) = \int_{y}^{\infty} Du; \quad Du \text{ is the Gaussian measure } Du = e^{-u^2/2}/\sqrt{2\pi}. \]

In terms of the fractional Hamming distance \( f = R/N \), \( r \) and \( s \) are defined as \( r = 1 - 2f \) and \( s = \sqrt{1 - r^2} \). Finally, the saddle point condition \( \frac{\partial G}{\partial q} = 0 \) becomes

\[
\frac{q}{2(1 - q)^2} + \alpha \int Dz D\tau \frac{\partial}{\partial q} \left( \int_{\tau}^{\infty} Dw H(y_0) \right) \int_{\tau}^{\infty} Dw H(y_0) = 0. \tag{6}
\]

This equation can be solved imposing the condition for replica symmetry, \( q \to 1 \), in which limit the entropy diverges to negative infinity as the fractional volume in the space of interactions shrinks to zero. In this limit, equation (6) reduces to

\[
\alpha_{\text{max}}^{-1} = \left[ \int_{-\infty}^{r x + s_1} D\tau \int_{-\kappa}^{\infty} Dz (z + \kappa)^2 + \frac{1}{r^2} \int_{x_1}^{\infty} D\tau \int_{-\kappa}^{\infty} Dz (r z + s + \kappa_1)^2 \right]. \tag{7}
\]

It can be checked from a standard stability analysis that this replica symmetric solution is in fact stable in the entire region of interest \( 0 < f < 1/2 \). The results of solving equation (7) numerically for \( \alpha \), for different values of \( f \), are shown in figure 1 for three different values of \( \kappa_1 \); we have set \( \kappa = 0 \) since this yields larger capacities. It can be seen from the figure that the capacity as a function of the radius of attraction \( f \) relative to the capacity at \( f = 0 \) decreases at a slower rate for larger values of \( \kappa_1 \). The maximum possible number of direct attractors corresponding to a given radius of direct attraction is given by the curve corresponding to \( \kappa_1 = 0 \).
In conclusion, we have calculated the maximum possible number of isotropic direct attractors of given radius of attraction in discrete state neural networks. This number depends on the 'strength' of the fixed point attractor, $\kappa_1$, is linear in the number of neurons $N$ and is a decreasing function of the radius of attraction; the rate of decrease is smaller for larger values of $\kappa_1$. There are several issues worthy of further investigation, of which we mention two. It would be interesting to determine optimal capacities of attractors which attract in two or more iterations. The construction of an algorithm which determines the weights of the network capable of maximal storage for a given radius of attraction would also be of considerable practical utility.

References