Universality in the space of interactions for network models

1989 J. Phys. A: Math. Gen. 22 2031

(http://iopscience.iop.org/0305-4470/22/12/008)

J. Phys. A: Math. Gen. **22** (1989) 2031-2038. Printed in the UK

Universality in the space of interactions for network models

L F Abbott and Thomas B Kepler

Physics Department, Brandeis University, Waltham, MA 02254, USA

Received 26 January 1989

Abstract. By modifying the measure used to sum over coupling matrices, we generalise Gardner's calculation of the fractional interaction-space volume and storage capacity of neural network models. We also compute the local field distribution for the network. The generalised measure allows us to consider networks with a wide variety of properties away from saturation, but we find that the original results for saturated networks are universal for all well behaved measures. Other universality classes including those containing Hebb matrices and pseudo-inverse matrices are obtained by considering singular measures.

1. Introduction

One of the most impressive and imaginative of Elizabeth Gardner's many contributions to neural network research was her analysis [1] of the space of interactions for network models. This pioneering work allows us to compute several important properties of neural network memories in a model-independent way. Consider an N-node network designed to store and recall αN uncorrelated patterns ξ_i^μ ($i = 1, \ldots, N$, $\mu = 1, \ldots, \alpha N$) using the couplings J_{ij}. Define

$$\gamma_i^\mu = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} J_{ij} \xi_i^\mu \xi_j^\mu . \tag{1.1}$$

Gardner computed the fractional volume in the space of all coupling matrices occupied by J_{ij} satisfying

$$\gamma_i^\mu > \kappa \tag{1.2}$$

and the normalisation condition

$$\sum_{j=1}^{N} J_{ij}^2 = N \tag{1.3}$$

for all i and μ. From this result she could determine the maximum storage capacity, $\alpha_c N$, as a function of κ:

$$\alpha_c = \left(\int_{-\kappa}^{\infty} \mathrm{D}z (\kappa + z)^2 \right)^{-1} \tag{1.4}$$

where we use the common notation

$$\mathrm{D}z = \mathrm{d}z \frac{\exp(-\frac{1}{2}z^2)}{\sqrt{2\pi}} . \tag{1.5}$$

In addition, using her techniques the distribution of γ values, $\rho(\gamma)$, in a network satisfying (1.2) and (1.3) can be computed [2]. Near saturation this distribution is

$$\rho(\gamma)=\frac{\exp(-\frac{1}{2}\gamma^2)}{\sqrt{2\pi}}\,\theta(\gamma-\kappa)+\delta(\gamma-\kappa)\int_{-\kappa}^{\infty}\mathrm{D}z. \tag{1.6}$$

Equations (1.4) and (1.6) are the main results of Gardner's program. Since they were produced using (1.2) and (1.3) it is natural to wonder about the effects of modifying these constraints. Modification of the normalisation constraint (1.3) has been considered [3]. Here we look at alternatives to the local field constraint (1.2). This will result in a change in the measure used to sum over coupling matrices in Gardner's approach. There are two reasons for considering such a modification. First, we might ask whether the results (1.4) and (1.6) for the capacity and local field distribution of a saturated network are in any way universal. In other words we will explore whether these results correspond to the universal critical behaviour of an entire class of models. We will show that in fact they do. Second, there are other limiting behaviours known for neural networks. For example, for a network of the Hebb type [4],

$$J_{ij}=\frac{1}{\sqrt{\alpha N}}\sum_{\mu=1}^{\alpha N}\xi_i^\mu\xi_j^\mu \tag{1.7}$$

storing αN patterns the γ distribution is given by

$$\rho(\gamma)=\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\gamma-\frac{1}{\sqrt{\alpha}}\right)^2\right] \tag{1.8}$$

and for the pseudo-inverse matrix [5]

$$J_{ij}=\frac{1}{\sqrt{\alpha(1-\alpha)N}}\sum_{\mu,\nu}C_{\mu\nu}^{-1}\xi_i^\mu\xi_j^\nu \tag{1.9}$$

with

$$C_{\mu\nu}=\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\xi_i^\nu \tag{1.10}$$

we have

$$\rho(\gamma)=\delta[\gamma-\sqrt{(1-\alpha)/\alpha}]. \tag{1.11}$$

We will see that these behaviours characterised by shifted Gaussian and δ-function distributions correspond to the limiting behaviour of other universality classes.

Our generalisation of the measure used for summing over coupling matrices is based on the following observations. Using Gardner's method we can compute the fractional volume occupied by J_{ij} satisfying (1.3) and the more specific constraint

$$\gamma_i^\mu=\Gamma_i^\mu \tag{1.12}$$

instead of the inequality (1.2). Here the Γ_i^μ are pre-assigned a fixed set of values. In this case the fractional volume of interaction space is given by

$$V_{\mathrm{T}}(\{\Gamma_i^\mu\})=\int\prod_{i\neq j}\mathrm{d}J_{ij}\prod_{\mu,i}\delta(\gamma_i^\mu-\Gamma_i^\mu)\prod_i\delta\left(\sum_{j\neq i}J_{ij}^2-N\right)\left[\int\prod_{i\neq j}\mathrm{d}J_{ij}\prod_i\delta\left(\sum_{j\neq i}J_{ij}^2-N\right)\right]^{-1} \tag{1.13}$$

where γ_i^μ is given by equation (1.1) and the δ functions impose the constraints (1.3) and (1.12). We now imagine that the Γ_i^μ are chosen stochastically from some *a priori* distribution $f(\Gamma)$. We choose a set of Γ_i^μ using this distribution, determine the fractional volume corresponding to that particular set and then repeat the procedure with another set of Γ_i^μ chosen from the same distribution, either summing or averaging the total interaction-space volumes. This will provide us with some measure of the difficulty of finding couplings J_{ij} starting with a given distribution $f(\Gamma)$ and will allow us to determine the storage capacity and resulting distribution $\rho(\gamma)$ obtained from such a procedure. It is important to carefully distinguish the variables Γ_i^μ which are the target values generated by the distribution $f(\Gamma)$ from the γ_i^μ which, as can be seen from equation (1.1), depend on the couplings J_{ij}. The distribution of γ_i^μ values is given by $\rho(\gamma)$ which is not the same as $f(\Gamma)$ because it is affected by the availability of couplings J_{ij} corresponding to a given Γ_i^μ set.

The average interaction-space fraction for a distribution of Γ_i^μ values given by $f(\Gamma)$ is just

$$\bar{V}_{\mathrm{T}} = \int \prod_{i,\mu} [\mathrm{d}\Gamma_i^\mu f(\Gamma_i^\mu)] V_{\mathrm{T}}(\{\Gamma_i^\mu\}). \tag{1.14}$$

We can compute this following the original calculation of Gardner [1]. In fact the original calculation just corresponds to the particular choice $f(\Gamma) = \theta(\Gamma - \kappa)$. The result, averaged over patterns ξ_i^μ, is

$$\left\langle \frac{1}{N} \ln \bar{V}_{\mathrm{T}} \right\rangle = G(q) \tag{1.15}$$

where

$$G(q) = \alpha \int \mathrm{d}z \frac{\exp[-(z^2/2q)]}{\sqrt{2\pi q}} \ln\left(\int \mathrm{d}x \frac{f(x) \exp[-(x-z)^2/2(1-q)]}{\sqrt{2\pi(1-q)}} \right)$$
$$+ \tfrac{1}{2} \ln(1-q) + \frac{q}{2(1-q)} \tag{1.16}$$

with q given by

$$\mathrm{d}G/\mathrm{d}q = 0. \tag{1.17}$$

The variable q is the typical overlap between two different coupling matrices both satisfying the constraints. Throughout we assume that replica symmetry is not broken. At saturation $q \to 1$ and we can determine the maximum capacity of the network by solving (1.17) for α with $q = 1$:

$$\alpha_{\mathrm{c}} = \left(\int \mathrm{d}z \int \mathrm{d}y (z-y)^2 A(z, y, 1) \right)^{-1} \tag{1.18}$$

where

$$A(z, y, q) = \frac{f(y) \exp[-(z-y)^2/2(1-q)] \exp[-z^2/2q]}{\sqrt{2\pi q} \int \mathrm{d}x f(x) \exp[-(x-z)^2/2(1-q)]}. \tag{1.19}$$

Following [2] we can also compute $\rho(\gamma)$ obtaining

$$\rho(\gamma) = \int \mathrm{d}z\, A(z, \gamma, q). \tag{1.20}$$

These results are a trivial extension of previous work. However, they will allow us to construct general classes of models and examine the universality of limiting behaviour near saturation.

2. General behaviour away from saturation

Away from saturation the modification in the measure which we have introduced allows us to obtain a wide range of γ distributions. For example near $q=0$ we find

$$\rho(\gamma)=\frac{f(\gamma)\exp(-\frac{1}{2}\gamma^2)}{\int \mathrm{d}x f(x)\exp(-\frac{1}{2}x^2)}. \tag{2.1}$$

This is just the *a priori* distribution cut off by a Gaussian factor reflecting the difficulty of finding coupling matrices with very large (or very small) γ values. To be more specific we can take a Gaussian *a priori* distribution

$$f(\Gamma)=\exp\left(-\frac{1}{2s^2}(\Gamma-t)^2\right). \tag{2.2}$$

Using the results in § 1 we find that for such a distribution

$$\rho(\gamma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(\gamma-\mu)^2\right) \tag{2.3}$$

with

$$\sigma=\frac{s\sqrt{1+[s^2/(1-q)^2]}}{1+[s^2/(1-q)]} \tag{2.4}$$

and

$$\mu=\frac{t}{1+[s^2/(1-q)]}. \tag{2.5}$$

The value of q for such networks is determined by

$$\alpha=\left[\left(\frac{2(1-\sigma^2)}{1+q+\sqrt{(1-q)^2+4q\sigma^2}}\right)^2+\frac{\mu^2}{q}\right]^{-1}. \tag{2.6}$$

These results indicate that by weighting the *a priori* Γ distribution and thus modifying the measure we can obtain a wide variety of behaviours arising from a broad range of underlying network models.

3. Universality near saturation

The variety of behaviours we found away from saturation disappears as we approach the saturation bound of the network. To see this we begin by considering distribution functions $f(\Gamma)$ which are bounded and differentiable in a region $\kappa<\Gamma<\kappa'$ and which vanish outside this region. We will compute the capacity and γ distribution near saturation. When $q\to 1$ we find from (1.19)

$$A(z,y,1)=\frac{\exp(\frac{1}{2}z^2)}{\sqrt{2\pi}}[\delta(z-y)\theta(y-\kappa)\theta(\kappa'-y)+\delta(y-\kappa)\theta(\kappa-z)+\delta(y-\kappa')\theta(z-\kappa')]. \tag{3.1}$$

Using this we find

$$\alpha_c=\left(\int_{-\kappa}^{\infty} Dz(\kappa+z)^2+\int_{\kappa'}^{\infty} Dz(z-\kappa')^2\right)^{-1} \tag{3.2}$$

and near saturation

$$\rho(\gamma)=\frac{\exp(-\frac{1}{2}\gamma^2)}{\sqrt{2\pi}}\,\theta(\gamma-\kappa)\theta(\kappa'-\gamma)+\delta(\gamma-\kappa)\int_{-\kappa}^{\infty} Dz+\delta(\gamma-\kappa')\int_{\kappa'}^{\infty} Dz. \tag{3.3}$$

Note that these results are independent of the function $f(\Gamma)$, only depending on the end points κ and κ'. In the limit $\kappa'\to\infty$ (that is, with no upper cutoff) these reduce to previous results [1, 2]. Thus, all distributions $f(\Gamma)$ which are sufficiently well behaved over a given interval produce identical networks near saturation. This is the universality class containing models constructed from the original Gardner constraint (1.2). It is easy to see that the example of the previous section falls in this class by taking the limit $q\to 1$ with s and t held fixed.

Because of the universality of the above results we must turn to more singular distributions $f(\Gamma)$ to explore other universality classes. To find different behaviour near saturation we need a distribution which is as singular as the exponential factors in $A(z, y, q)$ as $q\to 1$. Consider *a priori* distributions of the form (in this case we will not bother with an upper cutoff at κ' although it can easily be included)

$$f(\Gamma)=h(\Gamma)\theta(\Gamma-\kappa)\exp\left(-\frac{g(\Gamma)}{1-q}\right). \tag{3.4}$$

Since the variable q only appears in the course of the mean-field calculation it is important to clarify what we mean by having a q in the *a priori* distribution. We should really write the q which appears in the above equation as $\tilde{q}$, a free parameter. We can then compute the mean-field variable q from equation (1.17). It will be a function of both α and $\tilde{q}$. What we mean by writing a q in the above equation is that $\tilde{q}$ is chosen to satisfy the equation

$$q(\alpha,\tilde{q})=\tilde{q}. \tag{3.5}$$

With this understanding we will not bother distinguishing between $\tilde{q}$ and q. Using this *a priori* distribution we find that in the limit $q\to 1$ all the necessary integrals can be done by the saddle-point method and the result for any well behaved function h and any function g satisfying $1+g''>0$ is

$$A(z,y,1)=\frac{\exp(-\frac{1}{2}z^2)}{\sqrt{2\pi}}[(1+g''(y))\theta(y-\kappa)\delta(z-y-g'(y))+\delta(y-\kappa)\theta(\kappa+g'(\kappa)-z)] \tag{3.6}$$

where a prime denotes differentiation. Then

$$\rho(\gamma)=\frac{(1+g''(\gamma))\theta(\gamma-\kappa)}{\sqrt{2\pi}}\exp[-\tfrac{1}{2}(\gamma+g'(\gamma))^2]+\delta(\gamma-\kappa)\int_{-\infty}^{\kappa+g'(\kappa)} Dz \tag{3.7}$$

near the saturation point

$$\alpha_c=\left(\frac{1}{\sqrt{2\pi}}\int_{\kappa}^{\infty} dy(1+g''(y))(g'(y))^2\exp[-\tfrac{1}{2}(y+g'(y))^2]+\int_{-\infty}^{\kappa+g'(\kappa)} Dz(z-\kappa)^2\right)^{-1}. \tag{3.8}$$

Note that the well behaved function $h(\Gamma)$ does not enter into these formulae (although it certainly would affect the behaviour away from saturation) so they correspond to the limiting behaviour of entire universality classes.

An interesting special case is

$$g(\Gamma)=(1/2\sigma)(\Gamma-\mu)^2-\tfrac{1}{2}\Gamma^2 \tag{3.9}$$

giving a Gausian distribution near saturation

$$\rho(\gamma)=\theta(\gamma-\kappa)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(\gamma-\mu)^2\right)+\delta(\gamma-\kappa)\int_{(\mu-\kappa)/\sigma}^{\infty}\mathrm{D}z \tag{3.10}$$

with

$$\alpha_c=\left(\int_{(\kappa-\mu)/\sigma}^{\infty}\mathrm{D}z[\mu+(1-\sigma)z]^2+\int_{(\mu-\kappa)/\sigma}^{\infty}\mathrm{D}z(z+\kappa)^2\right)^{-1}. \tag{3.11}$$

Several limits of the above results are interesting. First, for $\mu=0$ and $\sigma=1$ we recover the original Gardner case (1.4) and (1.6). If we take $\sigma=1$ and $\kappa\to-\infty$ we find

$$\rho(\gamma)=\frac{1}{\sqrt{2\pi}}\exp[-\tfrac{1}{2}(\gamma-\mu)^2] \tag{3.12}$$

and

$$\alpha_c=1/\mu^2 \tag{3.13}$$

which is the result for a Hebb matrix (1.8) given in the introduction. Thus, the Hebb matrix represents the limiting behaviour of a whole class of models which have a shifted Gaussian γ distribution near saturation. The α_c given above is the maximum capacity at which the constraints we have imposed can be satisfied and is not related to the maximum capacity of the Hebb model [6]. When $\sigma\to 0$ and $\kappa<\mu$ we obtain

$$\rho(\gamma)=\delta(\gamma-\mu) \tag{3.14}$$

and

$$\alpha_c=1/(1+\mu^2) \tag{3.15}$$

which by comparison with (1.11) is the same behaviour as the pseudo-inverse model. Again the pseudo-inverse matrix represents the limiting behaviour of a whole other universality class of models.

The Hebb and pseudo-inverse cases do not exhaust the universality classes even for a Gaussian *a priori* distribution. Suppose for simplicity that we take $\kappa\to-\infty$ (no θ function). We are thus considering models like the Hebb model which make errors because nothing prevents some of the γ_i^μ from becoming negative. Then

$$\rho(\gamma)=\frac{1}{\sigma\sqrt{2\pi}}\exp[-(1/2\sigma^2)(\gamma-\mu)^2] \tag{3.16}$$

and

$$\alpha_c=1/[\mu^2+(1-\sigma)^2]. \tag{3.17}$$

Thus, the Hebb model is just the $\sigma=1$ case of a one-parameter family of universality classes. Note that when $\sigma=1$, α_c is unbounded as $\mu\to 0$. Of course this does not imply a memory with infinite capacity because we are not imposing stability on these patterns. We can compute the fraction of sites i which are unstable for arbitrary σ and μ by integrating the distribution function over all $\gamma < 0$:

$$F(\gamma < 0) = \int_{\beta}^{\infty} \mathrm{D}z \tag{3.18}$$

where

$$\beta = \mu/\sigma. \tag{3.19}$$

For the Hebb matrix at saturation $\beta = 2.67$ [6]. We can construct a matrix which performs as well as the Hebb matrix at its saturation point but which has a much larger capacity by holding β fixed at this value (and thus holding the fraction of unstable sites fixed) and maximising α_c as a function of σ which is now a free variable (while in the Hebb case it is fixed at $\sigma = 1$). Maximising

$$\alpha_c = 1/[\beta^2\sigma^2 + (1-\sigma)^2] \tag{3.20}$$

gives

$$\sigma = 1/(\beta^2 + 1) \tag{3.21}$$

and

$$\alpha_c = (\beta^2 + 1)/\beta^2 = 1.14 \tag{3.22}$$

a remarkably high value for a network with a purely Gaussian local field distribution performing as well as a Hebb network near saturation. We see that the Hebb network has a relatively small capacity because it has too broad a Gaussian distribution of γ values and that by narrowing this distribution using the *a priori* distribution $f(\Gamma)$ we can dramatically increase the network's capacity.

4. Conclusions

By generalising the initial assumptions of the Gardner approach we have shown that the distribution $\rho(\gamma)$ and capacity of a network near saturation fall into universality classes. For well behaved *a priori* distributions, $f(\Gamma)$, the limiting behaviour depends only on the boundaries of the support of $f(\Gamma)$ and is essentially that found using the original Gardner constraint. Other universality classes with the limiting behaviour given by (3.7) and (3.8) were obtained from more singular *a priori* distributions. It would be interesting to know what other sorts of limiting behavour besides that of (3.7) and (3.8) is possible. Algorithms exist [7] for constructing models in the Gardner universality class, and of course for constructing the Hebb and pseudo-inverse models. It may be possible to construct models exhibiting all of the universal behaviours we have seen by using a stochastic learning algorithm based on the initial distribution $f(\Gamma)$. Near saturation such models should exhibit the γ distributions and capacities which we have found.

Acknowledgment

This research was supported in part by US Department of Energy Contract DE-AC0276-ER03230.

References

[1] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[2] Kepler T and Abbott L F 1988 *J. Physique* **49** 1657
Krauth W, Nadal J-P and Mézard M 1988 *J. Phys. A: Math. Gen.* **21** 2995
[3] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
Krauth W 1988 *Preprint* Ecole Normale
[4] Hebb D C 1957 *The Organization of Behaviour: A Neurophysiological Theory* (New York: Wiley)
Cooper L N, Liberman and Oja E 1979 *Biol. Cybern.* **33** 9
Hopfield J J 1982 *Proc. Natl. Acad. Sci. USA* **79** 2254
[5] Personnaz L, Guyon I and Dreyfus G 1985 *J. Physique* **46** L359
Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[6] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
[7] Minsky M and Papert S 1969 *Perceptrons* (Cambridge MA: MIT Press)
Gardner E, Stroud N and Wallace D 1987 *Preprint* 87/394 Edinburgh
Poeppel G and Krey U 1987 *Europhys. Lett.* **4** 481
Diederich S and Opper M 1987 *Phys. Rev. Lett.* **58** 949
Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
Krauth W and Mézard M 1987 *J. Phys. A: Math. Gen.* **20** L745
Abbott L F and Kepler T 1988 Optimal learning in neural network memories *Preprint* Brandeis University BRX-TH-255