Critical capacity of noisy and asymmetrically constrained perceptrons

1991 J. Phys. A: Math. Gen. 24 1947

(http://iopscience.iop.org/0305-4470/24/8/032)

J. Phys. A: Math. Gen. 24 (1991) 1947–1958. Printed in the UK

Critical capacity of noisy and asymmetrically constrained perceptrons

Jorge Kurchan and Eytan Domany

Department of Electronics, Weizmann Institute of Science, Rehovot 76100, Israel

Received 19 September 1990

Abstract. We report calculations of the critical capacity of perceptrons that are subject to pattern-dependent noise. We also calculate the capacity of a perceptron whose weights take values on a shifted sphere, and show how this system resembles 'noisy' behaviour in some limits.

1. Introduction

In calculations of the average critical capacity of perceptrons, the couplings J_i are assumed to be constrained to take a set of possible values. In her original work [1], Gardner considered couplings restricted to take values on a sphere. Several other possibilities were later studied along the same lines, such as binary valued ($J_i = \pm 1$) [2, 3], spherical plus sign-constrained [4] and J_i taking limits of discrete set of values [5]. More recently we have presented a closed form geometrical expression for the capacity when the J_i are constrained to lie on a general surface [6].

In practical realizations of neural networks some of these constraints may appear naturally; for example, in an all-optical neural network if may be difficult to implement neurons with both excitatory and inhibitory incident connections [7].

The geometrical constraints mentioned above can, in principle, have a noisy component; for example, the allowed range of J_i may contain a part which varies randomly between sites i. In the case of a perceptron, such random variations of the constraints will have an effect on the output through generation of a 'synaptic noise', which is pattern dependent but otherwise fixed. Under certain circumstances this 'noise' can be considered to be uncorrelated to the output that corresponds to noise-free constraints.

In this paper we first study the effect of an uncorrelated pattern-dependent output noise on the capacity of simple perceptrons. Secondly we study the effect of a constant shift in the constraints satisfied by the J_i, and show that in the limit of large shifts the effect is equivalent to the previous noisy situation.

We follow the treatment of Gardner [1], who calculated $\alpha_c(k)$, the critical capacity for storing random patterns. The first problem we consider is one in which the embedding field of a pattern μ is subject to an (additive) noise t^μ. This noise may have any one of a variety of sources. For example, the single output unit may have a noisy threshold, $k \to k + t^\mu$. Alternatively, each of the N couplings J_j may be shifted away from its (spherically) constrained value by a non-constrained Gaussian distributed random term. The question we ask may be stated as follows: if the threshold k is

0305-4470/91/081947+12\$03.50

allowed to fluctuate, i.e. takes larger values when some patterns are presented and smaller values for others, how is the capacity affected?

Performing a quenched average for Gaussian noise with mean $\overline{t^\mu}=0$ and variance $\overline{(t^\mu)^2}=a^2$, we find that the critical capacity behaves as follows:

$$\alpha_c(a,k)=\frac{\alpha_c(0,k^{(a)})}{1+a^2} \tag{1}$$

where $\alpha_c(0,k)$ is the capacity calculated (with no noise, i.e. $a=0$) by Gardner, and

$$k^{(a)}=\frac{k}{(1+a^2)^{1/2}}. \tag{2}$$

The second problem studied is that of couplings J_j which do not have the same symmetry as the stored patterns, e.g., when $\boldsymbol{J}$ and $-\boldsymbol{J}$ occur with different weights, whereas ξ^μ and $-\xi^\mu$ have the same probabilities. We realize this situation by restricting the allowed vector $\boldsymbol{J}$ to lie on a sphere whose centre is shifted by $\delta J_j=a$ from the origin. This problem is a (rather instructive) particular case of a more general situation, in which the vectors $\boldsymbol{J}$ are constrained to any general hypersurface [6]. Finally we show that in the limit of large a the two problems become identical.

2. Brief review of Gardner's calculation

To introduce notation, we first summarize briefly the calculation of Gardner [1]. We consider a single output unit, whose state $S=\pm 1$ is determined according to the nonlinear rule

$$S=\operatorname{sign}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}J_jS_j-k\right). \tag{3}$$

The input units take values $S_j=\xi_j^\mu=\pm 1$, where $\mu=1,\ldots,p$ randomly generated input patterns are to be mapped onto $S=\xi^\mu=\pm 1$, also chosen at random. Gardner posed the following problem: what is the largest value of $\alpha=p/N$ for which J_i can be found, that satisfies p conditions on the embedding fields,

$$h^\mu=\xi^\mu\frac{1}{\sqrt{N}}\sum_j J_j\xi_j^\mu>k \tag{4}$$

under the constraint

$$\sum_{j=1}^{N}J_j^2=N. \tag{5}$$

The volume of weights in J-space that achieve correct mapping is given by

$$V(\{\xi^\mu\})=\int\prod_i \mathrm{d}J_i\,\delta\left(\sum_j J_j^2-N\right)\prod_\mu\theta(h^\mu-k). \tag{6}$$

Dependence on the patterns $\{\xi\}$ enters through h^μ, see (4). To obtain the typical volume, one has to take $\overline{\ln V}=(\mathrm{d}/\mathrm{d}n)\overline{V^n}|_{n\to 0^+}$, where the bar denotes averaging over the $\{\xi^\mu\}$ configurations.

Using the replica method, and expressing the δ and θ functions as integrals, $\overline{V^n}$ can be expressed as

$$\overline{V^n}=\int\prod_{\beta<\gamma}\frac{\mathrm{d}q_{\beta\gamma}\,\mathrm{d}\varphi_{\beta\gamma}}{2\pi}\prod_{\sigma}\mathrm{d}E_\sigma\exp\left[N\left(\alpha G_0(q_{\beta\gamma},k)+G_1(E_\beta,\varphi_{\beta\gamma})+\mathrm{i}\sum_{\beta<\gamma}\varphi_{\beta\gamma}q_{\beta\gamma}\right)\right] \tag{7}$$

with $\alpha=p/N$, and

$$\mathrm{e}^{G_0}=\prod_\beta\int_{\lambda_\beta=k}^{\infty}\mathrm{d}\lambda_\beta\int_{x=-\infty}^{\infty}\mathrm{d}x_\beta\exp\left(\mathrm{i}\sum_\gamma x_\gamma\lambda_\gamma-\sum_{\beta<\gamma}q_{\beta\gamma}x_\beta x_\gamma-\tfrac{1}{2}\sum_\beta x_\beta^2\right) \tag{8a}$$

$$\mathrm{e}^{G_1}=\int_{-\infty}^{\infty}\prod_\beta\mathrm{d}J_\beta\exp\left(\mathrm{i}\sum_\beta E_\beta(J_\beta^2-1)-\mathrm{i}\sum_{\beta<\gamma}\varphi_{\beta\gamma}J_\beta J_\gamma\right). \tag{8b}$$

The integral (7) is calculated by steepest descent. A replica symmetric minimum is assumed, of the form

$$q_{\alpha\beta}=q\qquad\varphi_{\alpha\beta}=\mathrm{i}F\qquad E_\alpha=\mathrm{i}E \tag{9}$$

with this simplification the expressions (8) reduce in the $n\to0$ limit to

$$G_0(q,k)=n\int_k^\infty\ln H[(\sqrt{q}\,t+k)/(1-q)^{1/2}]\,\mathrm{D}t \tag{10a}$$

$$H(x)=\int_x^\infty\mathrm{D}t,\qquad \mathrm{D}t=\frac{\mathrm{e}^{-t^2/2}\,\mathrm{d}t}{\sqrt{2\pi}} \tag{10b}$$

and

$$G_1(\mathrm{i}F,\mathrm{i}E)=n\left(\tfrac{1}{2}\ln2\pi-\tfrac{1}{2}\ln(2E+F)+\frac{1}{2}\frac{F}{2E+F}\right). \tag{11}$$

The remaining minimization problem yields

$$\frac{\ln V}{N}=\operatorname*{extr}_{E,F,q}\frac{1}{n}(\alpha G_0(q,k)+G_1(\mathrm{i}F,\mathrm{i}E)+\tfrac{1}{2}nFq). \tag{12}$$

As the number of patterns (and the corresponding conditions (4)) increases, the volume of solutions shrinks. When this volume shrinks to zero, the angle between all pairs of solutions approaches 0, and hence $q\to1$ as $\alpha\to\alpha_c$. The saddle point equations (12) have a solution in this limit only if the condition

$$\alpha<\alpha_c=\left(\int_{-k}^{\infty}\mathrm{D}t(t+k)^2\right)^{-1} \tag{13}$$

is satisfied. For $k=0$, $\alpha_c=2$, and $\alpha_c(k)$ decreases as k increases, $\alpha_c\sim1/k^2$ for large k.

3. Noisy patterns

We now turn to calculate the capacity of a perceptron whose embedding field is subjected to a Gaussian noise. First we state the problem we study, next we explain how it differs from other related calculations, and then we present the results. Consider a perceptron whose response to an input pattern is given by:

$$S=\operatorname{sign}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}J_iS_j+at-k\right) \tag{14}$$

Here the 'noise' is Gaussian distributed, $\overline{t^2}=1$ and $\bar{t}=0$ (note that the width a is now scaled out). As explained in the introduction, a noisy threshold will give rise to such an input-output relation. We perform a quenched average over the noise, which corresponds to the following physical situation: for each stored pattern μ, generate a shift at^μ of the threshold, and require that the relation

$$\xi^\mu = \mathrm{sign}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N} J_i S_j - (k - at^\mu)\right) \tag{15}$$

holds for every pattern. An alternative interpretation can be obtained by posing the following question: the couplings J_j contain a principal term J_j^*, which satisfy the spherical constraint $\Sigma_j (J_j^*)^2 = N$, and a random term δJ_j

$$J_j = J_j^* + \delta J_j. \tag{16}$$

A solution is found by searching the J_j^*-space, while keeping the δJ_j fixed. This 'frozen synaptic noise' adds to the field of the random patterns a Gaussian distributed noise of the form

$$at^\mu = \frac{1}{\sqrt{N}}\sum_j \delta J_j \xi_j^\mu \xi^\mu. \tag{17}$$

The effect of such 'static synaptic noise' were investigated for Hopfield networks [8] as well as for Hebbian feed-forward layered nets [9].

Here we calculate the volume of the space of interactions that will give the correct output to p patterns ξ_j^μ in the presence of the noise. One needs to perform a quenched average over the noise field. The volume of solutions is given by:

$$V(\{\xi^\mu\},\{t^\mu\}) = \int \prod_j \mathrm{d}J_j \delta\left(\sum_j J_j^2 - N\right) \prod_\mu \theta\left(\frac{1}{\sqrt{N}}\xi^\mu \sum_j J_j \xi_j^\mu + a\xi^\mu t^\mu - k\right) \tag{18}$$

and we evaluate

$$\overline{V^n} = \prod_{j,\mu}\left(\tfrac{1}{2}\sum_{\xi_j^\mu=\pm 1}\right)\left(\tfrac{1}{2}\sum_{\xi^\mu=\pm 1}\right)\left(\int \mathrm{D}t^\mu\right) V^n(\{\xi\},\{t\}) \tag{19}$$

where the bar now denotes averaging over the patterns $\{\xi^\mu\}$, and over the (quenched) noise field t^μ. Related problems are those of training in the presence of noise. These were addressed numerically [10] and analytically [11] for attractor networks, as well as for perceptrons [12–14].

When the noise is averaged for each set of selected patterns ξ, an annealed average over the noise is to be taken. In our case this corresponds to first integrating $V(\{\xi\},\{t\})$ over t, and only then averaging V^n over the patterns $\{\xi\}$.

It is interesting to see that had we considered an annealed average for the t^μ, the effect of the noise would have been to replace in the original Gardner's calculation the θ functions by the function:

$$\theta(x) \to f_a(x) = H\left(\frac{-x}{a}\right) \tag{20}$$

where H is the error function defined in (10b). One can compare this situation to the calculation of the partition function à la Gardner–Derrida [2] with inverse temperature β and the perceptron learning Hamiltonian

$$H = \sum_\mu \left(k - \sum_i \frac{J_i}{\sqrt{N}}\xi_i^\mu\right)\theta\left(k - \sum_i \frac{J_i}{\sqrt{N}}\xi_i^\mu\right). \tag{21}$$

In this case the θ functions of Gardner's calculation are replaced by the function:

$$f_\beta(x)=e^{\beta x}+(1-e^{\beta x})\theta(x). \tag{22}$$

Both are particular cases of the calculation of Abbott and Kepler [15]. The functions f_a and f_β are sigmoids of typical widths a and $1/\beta$ respectively.

Turning now to our quenched average over the noise field, note that the calculation of $\overline{V^n}$ proceeds along the same lines as Gardner's. We find that $\overline{V^n}$ is now given by an expression similar to (7), with the same $G_1(\{\varepsilon_\beta\},\{\varphi_{\beta\gamma}\})$ as (8*b*), but G_0 *replaced* by $G_{0,a}(\{q_{\beta\gamma}\},k)$, given by

$$e^{G_{0,a}}=\int_{\lambda_\beta=k}^{\infty}\int_{-\infty}^{\infty}\prod_\beta\frac{d\lambda_\beta}{2\pi}dx_\beta$$

$$\times\exp\left(i\sum_r x_\gamma\lambda_\gamma-\sum_{\beta<\gamma}q_{\beta\gamma}x_\beta x_\gamma-\tfrac{1}{2}\sum_\beta(x_\beta)^2-\tfrac{1}{2}a^2\sum_{\beta,\gamma}x_\beta x_\gamma\right). \tag{23}$$

This differs from (8*a*) only in the last term; obviously $G_{0,a=0}=G_0$ as expected, and the effect of $a\neq 0$ is easily shown (by changing variables in (23)) to be given by

$$G_{0,a}(\{q_{\beta r}\},k)=G_0(\{q^{(a)}_{\beta\gamma}\},k^{(a)}) \tag{24}$$

where

$$q^{(a)}_{\beta,\gamma}=\frac{q_{\beta,\gamma}+a^2}{1+a^2} \tag{25a}$$

and

$$k^{(a)}=\frac{k}{(1+a^2)^{1/2}}. \tag{25b}$$

Hence the entire effect of the quenched noise field is taken into account by properly rescaling the variables q, k. We proceed as before, assuming replica symmetry and minimizing the function

$$\frac{1}{N}\overline{\ln V}=\underset{E,F,q}{\text{extr}}\frac{1}{n}(\alpha G_0,(q^{(a)},k^{(a)})+G_1(iE,iF)+\tfrac{1}{2}nFq). \tag{26}$$

We are again interested in the $q\to 1$ limit. Note that

$$\frac{dq^a}{dq}=\frac{1}{1+a^2}. \tag{27}$$

Note also that if a function $f(q)$ diverges as $(1-q)^{-r}$ as $q\to 1$, then:

$$\lim_{q\to 1}[(1-q)^r f(q^{(a)})]=(1+a^2)^r\lim_{q\to 1}[(1-q)^r f(q)]. \tag{28}$$

Since dg_0/dq diverges as $(1-q)^{-2}$ as $q\to 1$ we have from (26)–(28) that in this limit the extremum equation is the same as in Gardner's calculation but with the substitution

$$\alpha_c\to\alpha_c(1+a^2) \tag{29}$$

and hence we find

$$\alpha_c(a,k)=\frac{\alpha_c(0,k^{(a)})}{1+a^2}. \tag{30}$$

Here the function $\alpha_c(0, k)$ is the critical capacity in the absence of noise, as given by (13). For $k=0$ we also have $k^{(a)}=0$, and hence

$$\alpha_c(a, 0)=\frac{2}{1+a^2} \tag{31}$$

and the capacity is degraded as $1/(1+a^2)$. On the other hand, in the limit of large $k^{(a)}$ one finds $\alpha_c(a, k) \simeq \alpha_c(0, k)$, as expected.

One can also easily show that the replica symmetric solution is stable with respect to transverse fluctuations if

$$\alpha_c \int_{-k^{(a)}}^{\infty} \mathrm{D}t - 1 < 0. \tag{32}$$

Inserting the expression for α_c this condition reads:

$$k^{(a)} \int_{-k^{(a)}}^{\infty} (t+k^{(a)})\,\mathrm{D}t + a^2 \int_{-k^{(a)}}^{\infty} (t+k^{(a)})^2\,\mathrm{D}t > 0. \tag{33}$$

We notice that for $a \neq 0$ this condition is satisfied even for some negative values of k.

4. Biased (shifted) J_i

We now turn to study the situation in which the weights J_j are constrained in a manner that violates the symmetry of the patterns ($\xi_j^\mu = \pm 1$ with equal probability). To be specific, we consider a spherical constraint of the form

$$\sum_j (J_j - a_j)^2 = N \tag{34}$$

where $a_j = \pm a$. The point $\boldsymbol{a}$ is the centre of a hypersphere in the space of weights, on which all J_j are restricted to lie. The volume of solutions is given by

$$V\{\xi_j\}\{a_j\}) = \int \prod_j \mathrm{d}\tilde{J}_j \delta\left(\sum \tilde{J}_j^2 - N\right) \prod_\mu \theta\left(\sum_j (\xi^\mu \tilde{J}_j \xi_j^\mu + \xi^\mu \xi_j^\mu a_j) - k\right) \tag{35}$$

where we changed variables to

$$\tilde{J}_j = J_j - a_j. \tag{36}$$

It is easy to see that under the 'gauge transformation'

$$\xi_j^\mu \to \xi_j^\mu \operatorname{sign} a_j \qquad \tilde{J}_j \to \tilde{J}_j \operatorname{sign} a_j$$

the problem becomes one in which all $a_j = +a$, i.e.

$$V(\{\xi\}, \{a_j\}) = V(\{\hat{\xi}\}, \{a_j = a\}).$$

As long as $\xi_j^\mu = \pm 1$ occur with equal probability, we can assume $a_j = +a$, without loss of generality. With this we can write (35) as

$$V(\{\xi\}, a) = \int \prod_j \mathrm{d}\tilde{J}_j\, \delta(\tilde{J}_j^2 - N) \prod_\mu (h^\mu + a t^\mu - k) \tag{37}$$

where we used

$$h^\mu = \frac{1}{\sqrt{N}} \sum_j \xi^\mu \tilde{J}_j \xi_j^\mu \tag{38}$$

$$t^\mu = \frac{1}{\sqrt{N}} \sum_j \xi^\mu \xi_j^\mu. \tag{39}$$

It is instructive to compare (37) with (14). The term at^μ, due to the shifted constraints, plays the same role as the quenched noise did in our previous calculation. Here, however, the 'noise' is correlated with the embedding field:

$$\overline{h^\mu t^\nu} = \delta_{\mu,\nu} \frac{1}{N} \sum_j \tilde{J}_j = \delta_{\mu,\nu} R_J. \tag{40}$$

When calculating $\overline{V^n}$, it is tempting to work with the full J_j variables; then the only change with respect to the Gardner calculation is in the argument of the spherical constraint, being now $\delta(\Pi_J(J_j - a)^2 - N)$. This temptation must, however, be resisted, since the corresponding order parameter $q_{\beta\gamma} = (1/N)\Sigma_j J_j^\beta J_j^\gamma$ is no longer the cosine of an angle, since $\boldsymbol{J}^\beta$ is not normalized! Therefore we lose the geometrical interpretation of q, and with it our advance knowledge that $q \to 1$ at α_c. Hence we prefer to work with the normalized variables $\tilde{J}_j$. With respect to the Gardner calculation one has now terms of the form

$$\sum_\mu \sum_{\beta<\gamma} x_\mu^\beta x_\mu^\gamma \left(\frac{1}{N} \sum_j J_j^\beta J_j^\gamma \right)$$
$$= \sum_\mu \sum_{\beta<\gamma} x_\mu^\beta x_\mu^\gamma \left(\left(\frac{1}{N} \sum_j \tilde{J}_j^\beta \tilde{J}_j^\gamma \right) + a \frac{1}{N} \left(\sum_j \tilde{J}_j^\beta + \sum_h \tilde{J}_j^\gamma \right) + a^2 \right). \tag{41}$$

Introducing the order parameters

$$q_{\beta\gamma} = \frac{1}{N} \sum_j \tilde{J}_j^\beta \tilde{J}_j^\gamma \qquad R_\beta = \frac{1}{N} \sum_j \tilde{J}_j^\beta \tag{42}$$

and their conjugate variables $\varphi_{\beta\gamma}$ and S_β, we can express

$$\overline{V^n} = \int \prod_{\beta<\gamma} \frac{\mathrm{d}q_{\beta\gamma}\, \mathrm{d}\varphi_{\beta\gamma}}{2\pi} \prod_\beta \frac{\mathrm{d}R_\beta\, \mathrm{d}S_\beta}{2\pi} \mathrm{d}E_\beta$$
$$\times \exp\left[N \left(G_{1R}(E_\beta, \varphi_{\beta\gamma}, S_\beta) + \alpha G_{0R}(q_{\beta\gamma}, R_\beta, K) \right.\right.$$
$$\left.\left. + \mathrm{i} \sum_{\beta<\gamma} \varphi_{\beta\gamma} q_{\beta\gamma} + \mathrm{i} \sum_\beta S_\beta R_\beta \right) \right] \tag{43}$$

where the functions G_{1R} and G_{0R} are given by

$$\mathrm{e}^{G_{1R}(E_\beta, \varphi_{\beta\gamma}, S_\beta)} = \prod_\beta \int \mathrm{d}J^\beta \exp\left(\mathrm{i} \sum_\beta E_\beta((J^\beta)^2 - 1) - \mathrm{i} \sum_{\beta<\gamma} \varphi_{\beta\gamma} J^\beta J^\gamma - \mathrm{i} \sum_\gamma S_\gamma J^\gamma \right) \tag{44}$$

$$\mathrm{e}^{G_{0R}(q_{\beta\gamma}, R_\beta, k)} = \prod_\beta \int_k^\infty \mathrm{d}\lambda_\beta \int_{-\infty}^\infty \mathrm{d}x_\beta \exp\left(\mathrm{i} \sum \prod_\beta x_\beta \lambda_\beta - \sum \prod_{\beta<\gamma} q_{\beta\gamma} x_\beta x_\gamma - \tfrac{1}{2} \sum_\beta x_\beta^2 \right.$$
$$\left. - \tfrac{1}{2} a^2 \sum_{\nu,\gamma} x_\beta x_\gamma - \tfrac{1}{2} a \sum_{\beta,\gamma} x_\beta x_\gamma (R_\beta + R_\gamma) \right). \tag{45}$$

This function can be rewritten as

$$\mathrm{e}^{G_{0R}(q_{\beta\gamma}, R_\beta, k)} = \prod_\beta \int_k^\infty \frac{\mathrm{d}\lambda\beta}{2\pi} \int_{-\infty}^\infty \mathrm{d}x_\beta \exp\left\{ \mathrm{i} \sum_\beta x_\beta \lambda_\beta - \tfrac{1}{2} \sum_\beta (1 + a^2 + 2aR_\beta) x_\beta^2 \right.$$
$$\left. - \sum_{\beta<\gamma} (q_{\beta\gamma} + a(R_\beta + R_\gamma) + a^2) x_\beta x_\gamma \right\}. \tag{46}$$

Changing variables, and introducing

$$k_\beta^{(R)}=k/(1+a^2+2aR_\beta)^{1/2}$$
$$q_{\beta\gamma}^{(R)}=\frac{q_{\beta\gamma}+a(R_\beta+R_\gamma)+a^2}{(1+a^2+2aR_\beta)^{1/2}(1+a^2+2aR_\gamma)^{1/2}} \tag{47}$$

the expression for $G_{0R}(q_{\beta\gamma}, R_\beta, k)$ takes the form of (7). In particular, with the replica symmetric ansatz of the previous sections, supplemented by

$$R_\beta=R \qquad S_\beta=\mathrm{i}S$$

we find that again, G_{0R} can be expressed in terms of the Gardner function G_0, with rescaled variables:

$$G_{0R}(q, R, k)=G_0(q^{(R)}, k^{(R)}) \tag{48}$$

where

$$q^{(R)}=\frac{q+2aR+a^2}{1+a^2+2aR} \tag{49a}$$

$$k^{(R)}=\frac{k}{(1+2aR+a^2)^{1/2}}. \tag{49b}$$

As to the other function, G_{1R}, it is easy to show that

$$G_{1R}(\mathrm{i}E, \mathrm{i}F, \mathrm{i}S)=G_1(\mathrm{i}E, \mathrm{i}F)+\frac{nS^2}{2(2E+F)} \tag{50}$$

where, again, $G_1(\mathrm{i}E, \mathrm{i}F)$ is the Gardner function (11). Therefore $\overline{V^n}$ is calculated, as before, by steepest descents and we obtain

$$\frac{1}{N}\overline{\ln V}\ \underset{q,R,S,E,F}{\mathrm{extr}}\ \frac{1}{n}\left(\alpha G_0(q^{(R)}, k^{(R)})+G_1(\mathrm{i}E, \mathrm{i}F)+\frac{ns^2}{2(2E+F)}+\tfrac{1}{2}nF_q+nSR\right). \tag{51}$$

These extremum conditions can be solved for F, E and S in terms of R and q:

$$F=\frac{q-R^2}{(1-q)^2} \tag{52a}$$

$$E=\frac{1-2q+R^2}{2(1-q)^2} \tag{52b}$$

$$S=\frac{R}{1-q}. \tag{52c}$$

The remaining equations are

$$\frac{\alpha}{n}\frac{\partial G_0}{\partial R}(q^{(R)}, k^{(R)})=\frac{R}{1-q} \tag{53a}$$

$$\frac{\alpha}{n}\frac{\partial}{\partial q}G_0(q^{(R)}, k^{(R)})=-\frac{q-R^2}{2(1-q)^2}. \tag{53b}$$

These again simplify considerably in the $q\to 1$ limit, for which they reduce to two equations, for α_c and for R, of the form

$$\frac{1-R^2}{1+2aR+a^2}=\alpha_c\int_{-k^{(R)}}^{\infty}\mathrm{D}t(t+k^{(R)})^2 \tag{54a}$$

$$-\frac{R}{a}=\alpha_c\int_{-k^{(R)}}^{\infty}\mathrm{D}t(t+k^{(R)})t. \tag{54b}$$

For $k=k^{(R)}=0$ the resulting quadratic equation has two solutions;

$$R=-a \qquad R=-\frac{1}{a}. \tag{55}$$

Since we must have, by definition, (see (42)) $|R|\leqslant 1$, the only acceptable solution has the form

$$R=\begin{cases}-a & |a|\leqslant 1\\ -1/a & |a|\geqslant 1\end{cases} \tag{56}$$

which upon substitution in (54) yields immediately the critical capacity

$$\alpha_c=\begin{cases}2 & |a|\leqslant 1\\ 2/a^2 & |a|\geqslant 1.\end{cases} \tag{57}$$

The result given in (57) for $k=0$ has a simple geometrical interpretation, which will be presented below. For finite k one can solve (53) only numerically. However, in the small k limit, one can expand and obtain

$$\frac{\alpha_c(k)}{\alpha_c(0)}=1-\frac{8}{\pi}\frac{k}{|1-a^2|^{1/2}}+O(k^2) \tag{58}$$

we present $\alpha_c(a)$ for various k in figure 1, and the numerical results for R in terms of a in figure 2.

Clearly, when a is large and $-R$ is small the capacity tends to the 'noisy' result (30); and the effect of shifted spherical constraint on α_c is the same as that of quenched noise. (Note that when $-R$ is small the 'shift term' (39) becomes typically uncorrelated with the embedding field cf (40).)

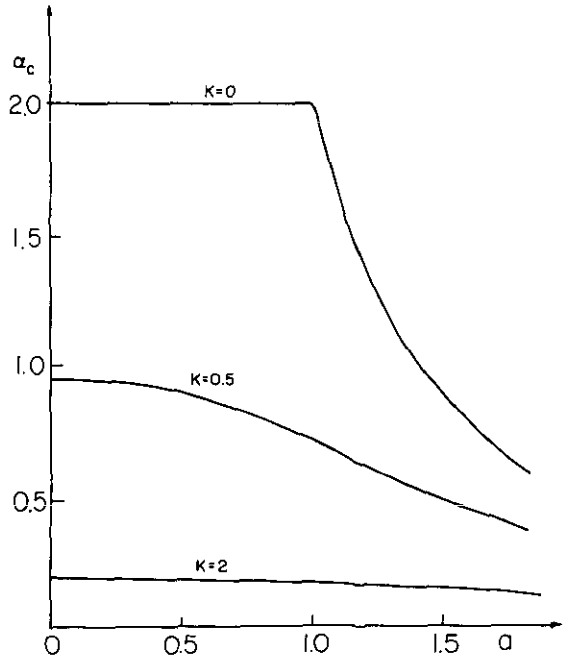

Figure 1. Critical capacity in terms of the shift parameter a.

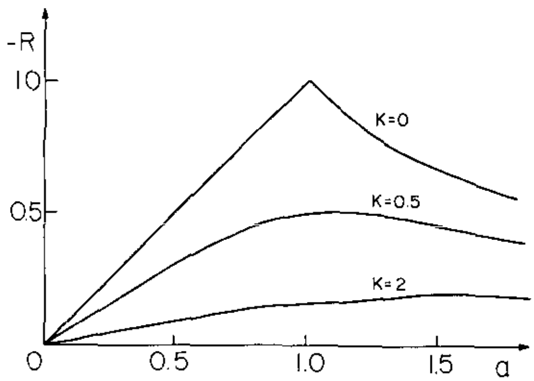

Figure 2. Value of the parameter R in terms of a.

Finally we have checked the stability of the replica-symmetric solution and found that it is stable against transversal fluctuations provided

$$\int_{-k^{(R)}}^{\infty} \mathrm{D}t\left(1+\frac{a}{R}\right)<0. \tag{59}$$

For $|a|<1$ the solution is stable for $k>0$, as in Gardner's case. For $|a|>1$, however, the limit of stability is shifted to negative values: the solution is found to be stable for $k>k^s<0$, precisely as was the case for noisy patterns.

The previous result has a simple geometric interpretation. In figure 3 we show the sphere of values of J_i (note that for $a<1$ the origin is inside and for $a>1$ outside it). Over it we have drawn two points J_i^β, J_i^γ corresponding to two replicas. We have in figure 3

$$\cos\alpha = q_{\beta\gamma} \qquad \cos\alpha^R = q_{\beta\gamma}^R \tag{60a}$$

and also, for $\lambda=\beta,\gamma$

$$\cos\theta_\lambda = R_\lambda \qquad d_\lambda = 1+a^2+2aR_\lambda \tag{60b}$$

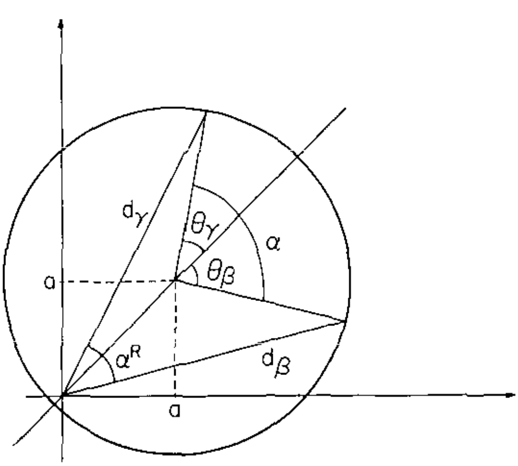

Figure 3. Geometric meaning of various quantities (see text).

For $k=0$ and a given set of patterns ξ_i^μ the space of solutions is the intersection of the set defined by the θ-function in (35) (which has the shape of a pyramid with centre in the origin) and the sphere defined in (34). The number q^R turns out to be the typical cosine of the angle between two *solutions* (from the origin) and its square root is roughly the cosine of the typical angle from the centre to some edge of the set.

For $q \to 1$, $q^R \to 1$ this angle tends to zero. Since the position set of solutions is distributed for different sets of patterns isotropically from the origin† we expect that most solutions will lie on a region of the sphere which subtends the largest solid angle from the origin. Considering annular domains on the sphere (34) with fixed values of R, we note that the solid angle subtended by them is proportional in the large-N limit to a sharply peaked function, corresponding to the magnitude (neglecting smooth functions of R)

$$\delta\Omega \sim \left(\frac{\sin\theta}{d}\right)^N = \left(\frac{1-R^2}{1+a^2+2aR}\right)^{N/2}. \tag{61}$$

By saddle point evaluation of (61) we recover the values (55). In figure 4 we show the predominant angles for $a<1$ and $a>1$. It is interesting to note that for $a>1$ the domain of solutions is two disconnected sets. It is only connected for the particular value of R that predominates in the large N limit (cf figure 4). As to the actual value of the critical capacity, it can also be understood in geometrical terms, considering the fact that for $a>1$ some solutions which are present in the $a<1$ case are lost. The geometric dependance of the capacity of perceptrons whose weights are constrained in general ways is discussed elsewhere [6].

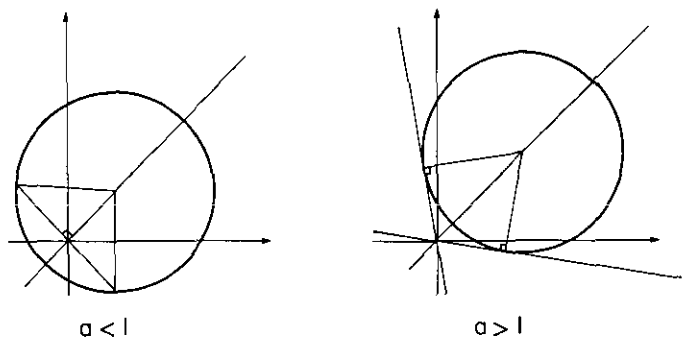

Figure 4. The regions of solutions predominating for large N.

JK thanks the Campomar-Weizmann fund for a fellowship. ED thanks Minerva for partial support.

References

[1] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[2] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271

† This is the case even if the distribution of the vector ξ_i^M is not isotropical as is the case for $\xi_i^M = \pm 1$, because angle correlations are lost through large numbers of patterns.

[3] Krauth W and Mezard M 1990 *Preprint*
[4] Amit D J, Wong K Y M and Campbell C 1989 *J. Phys. A: Math. Gen.* **22** 2039
[5] Gutfreund H and Stein Y 1990 *J. Phys. A: Math. Gen.* **23** 2613
[6] Kurchan J and Domany E 1990 *J. Phys. A: Math. Gen.* **23** L847
[7] Shouval H, Shariv I, Grossman T, Friesem A A and Domany E *Int. J. Neural Networks* in press
[8] Sompolinski H 1987 *Proc. Heidelberg Colloq. on Glassy Dynamics of Optimization (June 1986)* (Berlin: Springer)
[9] Meir R 1989 *J. Physique* **49** 201
[10] Gardner E, Stroud N and Wallace W J 1989 *J. Phys. A: Math. Gen.* **22** 2019
[11] Wong K Y M and Sherrington D 1990 *J. Phys. A: Math. Gen.* **23** L175; 1990 *Europhys. Lett.* **10** 419
[12] Refregier Ph and Vignolle J M 1989 *Europhys. Lett.* **10** 387
[13] Hansel D and Sompolinski H 1990 *Europhys. Lett.* **11** 687
[14] Gyorgyi G and Tishby N 1989 *Proc. STATPHYS 17*
[15] Abbott L F and Kepler T B 1989 *J. Phys. A: Math. Gen.* **22** 2031