

Prototype extraction in material attractor neural network with stochastic dynamic learning

S. Fusi

INFN, Sezione dell'Istituto Superiore di Sanità
Viale Regina Elena 299, I-00185 Roma

ABSTRACT

Dynamic learning of random stimuli can be described as a random walk among the stable synaptic values. It is shown that prototype extraction can take place in material attractor neural networks when the stimuli are correlated and hierarchically organized. The network learns a set of attractors representing the prototypes in a completely unsupervised fashion and is able to modify its attractors when the input statistics change. Learning and forgetting rates are computed.

Keywords: attractor neural networks, stochastic learning

1 INTRODUCTION

Unsupervised learning of uncorrelated patterns in attractor networks was recently described¹ as a stochastic process on the distribution of synaptic values characterizing the network. In this approach synapses have a finite number of stable states (efficacies) and learning is schematized as a random walk among them. Probability of each step is determined by the activities of the two neurons connected by the synapse. Neuronal states are in turn driven by external stimuli, represented as a stream of words, or patterns, each consisting of N 0's or 1's. Patterns have a given coding level f , which is the average fraction of neurons driven by a stimulus.

Stimuli drive synaptic transitions by means of the activity correlations of the two neurons connected by each synapse. If both neurons on a synapse have high activity there is a non-zero probability for the synapse to transit into an higher efficacy state; anticorrelation produces transitions depressing the synaptic efficacies; two inactive neurons leave the synapse unchanged.

Synaptic modifications induced by more recently presented patterns stack upon those produced by older patterns (*palimpsestic* property^{6,5}). As more patterns are presented, traces of older ones gradually fade away. After the learning of a given pattern, the network can learn other patterns without destroying the retrieval of the first one. Thus, the network reaches and remain into a steady state, with new patterns learned starting from an *asymptotic distribution* of synaptic weights.

Always the same number of patterns is stored in memory, ready for retrieval. Retrievable patterns are the most recently learned ones. This limit has been investigated^{1,2} by checking the bitwise stability of a recalled pattern by a signal-to-noise computation. In other words, after the learning of p uncorrelated patterns in the stream, the first pattern is presented to the network for retrieval. The synaptic input generated by this pattern, given the resulting learned synaptic matrix, is checked to see if it reproduces the originally learned pattern. It turns out that, with fixed learning parameters and a stream of uncorrelated patterns, a network cannot store any patterns without errors. Allowing the coding level f and the transition probabilities to vary appropriately with the number of neurons N as $N \rightarrow \infty$, one can achieve optimal storage and retrieve as many as $\sim (N/\log N)^2$ patterns.

If in the stream of stimuli there are groups of similar, highly overlapping patterns, with larger "distances"

(e.g. in the sense of Hamming) separating patterns of different groups, one would like the resulting attractors to create a relatively stable representative for each group (the prototype). The learning process should be able to extract the prototype and to create a synaptic structure more correlated to the prototype than to each one of the members belonging to the same class, even if the prototype is never presented to the network. This possibility is related to:

1. the *structure of the patterns*: if the correlation among different stimuli belonging to the same group is too low, prototype extraction cannot take place. In this case even speaking of a “class” does not make sense.
2. the *learning speed*: the structure of a prototype attractor can depend only on the patterns whose trace is still in the synaptic structure. If learning is fast, erasure of the traces of previous patterns is fast and only the traces of a small number of patterns are available. If many patterns of a given group appear together, the corresponding attractor for the group will have traces of them all. Though the last presented pattern will have a very pronounced weight. On the other extreme, if the patterns from different groups of similar patterns are intermixed in the input stream and only one pattern of each group appears in the memory, then the structure of the attractor will reflect this pattern only. It will still represent the group, in the sense that stimuli similar to it will flow into it. But, this representative will be very unstable. On the contrary if learning is slow, the memory span is long and the change in the synaptic structure induced by each stimulus is small. The attractor representing each group will be a mixture of all the patterns of the same group whose memory is still preserved in the synaptic structure. The attractor will be some mean of the stimulus group, and it can be considered a good prototype for all the patterns of the group. After each presentation of a new pattern of the same group, the attractor will move very slowly toward this new pattern preserving the information about all the other stimuli belonging to the same group which had been learnt previously. The speed with which an attractor will follow new samples from the same group will depend on the number of patterns of the same group which are held in memory, that, in turn, depends on the speed of learning.
3. the *number of different classes* presented to the network: if the network is presented a large number of different groups, the memory span cover only a small number of stimuli belonging to each class.

We shall find how these factors are related to the parameters of the network and in which regions prototype extraction can take place. In the first part of the paper we pass in review the main points of the theoretical framework. Then we shall study the learning process in the case of multiple presentations of the same pattern, and finally we will find some general constraints regarding the prototype extraction.

1.1 The theoretical framework

The framework is the same as that of reference¹: the network is presented a stream of p random stimuli. Since the synapses are assumed to have a finite number of states, after the removal of each stimulus, the synaptic efficacies converge rapidly to one of the stable states. So, on long time scales, the learning process can be seen as a walk among the stable states. In fact this is a random walk for the following two reasons:

1. the stimuli are assumed random, so each synapse will see a random sequence of pairs of activities (ξ_i, ξ_j) on the two neurons connected by it^{1,4,8}
2. The transition from one stable synaptic state to another may itself be stochastic.^{1,4,3} There are at least two possible sources of synaptic stochasticity: the dynamics of the synaptic modification and the random fluctuations in the spike rates of the two neurons connected by the synapse.

This stochastic process is described by a conditional probability distribution function for the synaptic values $\rho_J^p(\xi, \tilde{\xi})$, giving the probability of finding the synaptic value J following the presentation of p uncorrelated patterns, conditional on the appearance of neural activities $\xi, \tilde{\xi}$ in the presentation of the first of the p patterns. Since the synapses are assumed to have a finite number of states, the stochastic process is generically ergodic⁹ and leads to an asymptotic distribution after a large enough number of presentations. This distribution is taken to be the initial distribution upon which learning takes place.

Also the conditional distribution $\rho_J^p(\xi, \tilde{\xi})$ is driven towards the asymptotic form, as $p \rightarrow \infty$, by the presentation of p uncorrelated patterns following the first one, the one that produced the conditioning. The dynamics of the

conditional distribution is

$$\rho_J^p(\xi, \tilde{\xi}) = \sum_K \rho_K^1(\xi, \tilde{\xi})(M^{p-1})_{KJ}, \quad (1)$$

where M is the stochastic transition matrix of the process: M_{KJ} is the probability of going from state K to state J . For instance consider the model¹ in which the network is presented sparse patterns (f , the mean fraction of active neurons, is small), the synapse has two stable states, and both depression and potentiation are structured:

$$M = \begin{pmatrix} 1 - f(1-f)q_- & f(1-f)q_- \\ f^2q_+ & 1 - f^2q_+ \end{pmatrix} \quad (2)$$

f^2q_+ is the probability of potentiation (the probability of having two active neurons produces the f^2 factor), and $f(1-f)q_-$ is the probability of depression.

1.2 Observables

The next step is to introduce some observables which measures the correlation between the synaptic structure and the stimuli. These quantities should enable us to study some fundamental aspects of the learning process without specifying the details of neural dynamics. Obviously the conditions for having prototype extraction are necessary and not sufficient. Only by specifying the dynamics of the neural activity we could go a step further and find sufficient conditions.

In what follows we shall compute the average synaptic value for connections inside a population of neurons active in pattern μ^{11} :

$$\rho_\mu = \frac{1}{fN(fN-1)} \sum_{i,j} J_{ij} \xi_i^\mu \xi_j^\mu$$

This quantity measures how well a particular memory has been learned. If the synapse has two stable states ($J_{ij} = 0, 1$), it can be estimated by:

$$\rho = \Pr(J_{ij} = 1 | \xi_i^\mu = 1, \xi_j^\mu = 1)$$

For a synaptic matrix not correlated to pattern ξ_i^μ , ρ_μ is equal to the asymptotic connectivity c . If all the connections between the foreground neurons have been potentiated, then $\rho_\mu = 1$. We will refer to this observable as intra-class connectivity (ICC).¹¹

It can be proved that in a model of realistic analog neurons¹¹ the ICC contains most of the information concerning the retrievability, at least for uncorrelated patterns and low loading level. In that case, if the ICC is below a threshold value, the reverberations are not able to maintain themselves and the attractor corresponding to spontaneous activity prevails. Above the threshold the attractor becomes stable. Also in the case of correlated patterns, the ICC gives valuable informations^{11,12} provided that the number of classes is far below the critical storage capacity.

2 MULTIPLE PRESENTATIONS OF THE SAME PATTERN

As a first step consider the case in which a network whose synapses have two stable states is presented the following stream of sparse stimuli:

$$\xi_i^1, \xi_i^2, \dots, \xi_i^{p+1}, \xi_i^1, \xi_i^{p+2}, \dots, \xi_i^{2p+1}, \dots, \xi_i^1, \xi_i^{np+2}, \dots, \xi_i^{(n+1)p+1}$$

Pattern ξ_i^1 is presented n times and p random uncorrelated patterns (ξ_i^μ , with $\mu \geq 2$) interpose between two successive presentations. So in the stream we have n ‘‘fully correlated’’ patterns (the n occurrences of ξ_i^1). We now proceed to compute the ICC of ξ_i^1 .

First the network is presented the stimulus ξ_i^1 . The conditional stochastic transition matrix $M(\xi_i^1 = 1, \xi_j^1 = 1)$ is

$$M(1, 1) = \begin{pmatrix} 1 & 0 \\ q_+ & 1 - q_+ \end{pmatrix}$$

The only source of randomness is in the intrinsic stochastic behaviour of the synapse. The value of the activities that the two neurons connected by the synapse had during the presentation of the stimulus are fixed: both of them were active. Following the first presentation of the pattern, the network is presented a series of p uncorrelated patterns $\xi^2, \xi^3, \dots, \xi^{p+1}$. Each of them changes the conditional probability distribution with the following stochastic matrix

$$M = \begin{pmatrix} 1 - 2q_-f(1-f) & 2q_-f(1-f) \\ q_+f^2 & 1 - q_+f^2 \end{pmatrix} \quad (3)$$

The total change in the conditional distribution can be computed by multiplying by M^p . Thus, after the first presentation of the stimulus, followed by p uncorrelated patterns, the stochastic matrix is:

$$M(1,1)M^p = \begin{pmatrix} 1 - A & A \\ q_+(1 - A) + (1 - q_+)B & 1 - q_+(1 - A) + (1 - q_+)B \end{pmatrix}$$

where $A = (1 - c)(1 - \lambda^p)$, $B = c(1 - \lambda^p)$, c is the asymptotic distribution $c = q_+f^2/(q_+f^2 + 2q_-f(1 - f))$ and λ is the eigenvalue of 3: $\lambda = 1 - q_+f^2 - q_-f(1 - f)$.

Then, stimulus ξ^1 is presented again to the network (second presentation), followed by the p uncorrelated stimuli. After n repetitions of the same presentation procedure (stimulus, p uncorrelated patterns), the total stochastic transition matrix is $(M(1,1)M^p)^n$ where n is the number of presentations of ξ_i^1 . Starting from the asymptotic distribution $(c, 1 - c)$, the final ICC of ξ_i^1 is given by the first component of the distribution vector $(\rho_{mp}, 1 - \rho_{mp}) = (c, 1 - c)(M(1,1)M^p)^n$:

$$\rho_{mp} = c\mu^n + c_{mp}(1 - \mu^n)$$

where $\mu = (1 - q_+)\lambda^p$ and c_{mp} , the asymptotic ICC ($n \rightarrow \infty$), holds

$$c_{mp} = \frac{q_+\lambda_M^p + c(1 - \lambda_M^p)}{1 - (1 - q_+)\lambda_M^p}$$

2.1 Learning and forgetting rates

During the learning phase the ICC tends exponentially to the asymptotic value c_{mp} :

$$\rho_{mp} = c\mu^n + c_{mp}(1 - \mu^n)$$

The speed of this approaching depends on μ : the smaller μ , the higher the approaching speed. If $q_+ = 1$ (deterministic up transitions) then the ICC does not increase after multiple presentations of the same pattern. As a consequence the stimulus must be learnt in a single shot.

If now we stop presenting stimulus ξ^1 and the network is stimulated by uncorrelated patterns, the synaptic structure starts forgetting ξ^1 . The forgetting rate is given by the expression of the ICC of stimulus ξ^1 when the stimulus ξ^1 is no more presented to network:

$$\rho_{mp} = \rho_0\lambda^{n'(p+1)} + c(1 - \lambda^{n'(p+1)})$$

where ρ_0 is the ICC of ξ_i^1 after n presentations, and n' is the number of groups of $p + 1$ uncorrelated patterns. So, the ‘‘time’’ constant for learning can be estimated by $\tau_l = \mu = (1 - q_+)\lambda^p$, while that of forgetting is $\tau_f = \lambda^{p+1}$. As a consequence, we have:

$$\frac{\tau_f}{\tau_l} = \frac{\lambda}{1 - q_+} = \frac{1 - q_+f^2 - 2q_-f(1 - f)}{1 - q_+}$$

If the coding level is low ($f \ll 1$), then

$$\frac{\tau_f}{\tau_l} \simeq \frac{1}{1 - q_+}$$

This means that, if q_+ is small, the learning rate can be much higher than the forgetting rate. To get a faster forgetting, the network must be presented high coding level patterns.

3 GENERALIZATION CAPABILITY

3.1 Statistics of hierarchically correlated stimuli: father and sons

We use three different kinds of stimuli: the father, his sons and the uncorrelated patterns. The father η_i^1 is the prototype of a class of similar stimuli (the sons). His N bits are chosen independently, at random, according to the following distribution:

$$\Pr(\eta_i^1 = 1) = f, \quad \Pr(\eta_i^1 = 0) = 1 - f$$

The average number of active neurons ($\eta_i = 1$) is fN . The coding level f is typically low¹¹ ($f = 0.05 - 0.10$).

The bits of the members $\eta_i^{1\mu}$ (the sons) belonging to the same class of η_i^1 , are generated from the bits of the father according to the following rule:

- If the i -th bit of the father is 1 then, with probability u , the son has an active i -th bit:

$$\Pr(\eta_i^{1\mu} = 1 | \eta_i^1 = 1) = 1 - (1 - f)(1 - m) = u$$

- Else the son has a zero bit with probability:

$$\Pr(\eta_i^{1\mu} = 0 | \eta_i^1 = 0) = 1 - f(1 - m) = v$$

The probabilities u and v are chosen in such a way that the sons have the same average number of active neurons as the father. m is the parameter which measures the similarity between the father and the sons: if $m = 1$ the father is identical to his descendants; if $m = 0$ the father is not correlated with his sons.

Consider now the following stream of stimuli:

$$\eta_i^{\mu 1}, \xi_i^1, \dots, \xi_i^p, \eta_i^{\mu 2}, \xi_i^{p+1}, \dots, \xi_i^{2p}, \dots, \eta_i^{\mu n}, \xi_i^{np+1}, \dots, \xi_i^{(n+1)p}$$

The network is presented n different sons interposed by groups of p uncorrelated random patterns. The father is never presented but it is correlated to the sons.

3.2 Father's ICC

The conditional stochastic transition matrix for the father, upon the presentation of a son, is given by:

$$M(\eta_i^1 = 1, \eta_j^1 = 1) = \begin{pmatrix} 1 - q_{f-} & q_{f-} \\ q_{f+} & 1 - q_{f+} \end{pmatrix}$$

where $q_{f+} = q_+ u^2$ is the probability that the synaptic efficacy goes from 0 to 1 when the father has $\eta_i^1 = 1$, $\eta_j^1 = 1$. The extra factor u^2 is the probability that the synapse sees a couple of active neurons when a son is presented and the father have had two active neurons. Note that when the network is presented a son and the two neurons of the father are active, there is also a non null probability of going down. In fact it may be the case that $\eta_i^{1\mu} = 1 = \eta_i^1 = 1, \eta_j^{1\mu} = 0 \neq \eta_j^1 = 1$ or $\eta_i^{1\mu} = 1 = \eta_j^1 = 1, \eta_i^{1\mu} = 0 \neq \eta_i^1 = 1$. So the conditional probability of a down transition is $q_{f-} = 2q_-(1 - u)u$.

Following the presentation of n sons, alternated by sequences of p uncorrelated patterns, the ICC of the father can be computed along the lines of the previous section:

$$\rho_f = c\mu_f^n + c_f(1 - \mu_f^n)$$

where $\mu_f = (1 - 2q_-(1 - u)u - q_+ u^2)\lambda^p$, and the asymptotic ICC, c_f , is:

$$c_f = \frac{q_+ u^2 \lambda^p + c(1 - \lambda^p)}{1 - \mu_f}$$

3.3 Sons'ICC

The correlation among different brothers are weaker than those between father and sons. In fact we have:

$$\Pr(\eta^{1\mu} = 1 | \eta^{1\nu} = 1) = 1 - (1 - f)(1 - m^2) = U$$

where μ and ν are the indexes of two different brothers.

The ICC of the firstborn (the first son presented to the network), following the presentation of $M-1$ brothers, is:

$$\rho_s = c\mu_s^M + c_s(1 - \mu_s^M)$$

where $\mu_s = (1 - 2q_-(1 - U)U - q_+U^2)\lambda^p$, and

$$c_s = \frac{q_+U^2\lambda^p + c(1 - \lambda^p)}{1 - \mu_f}$$

3.4 Father vs sons

After the presentation of the firstborn, the ICC of the father is lower than that of his sons and the ICC of the future brothers (not yet presented) are even lower. In fact, following the presentation of the firstborn and p uncorrelated patterns, we have:

$$\rho_f = c + (1 - c)q_+u^2\lambda^p - 2cq_-u(1 - u)\lambda^p$$

$$\rho_s = c + (1 - c)q_+\lambda^p$$

The ICC of the son is always greater than that of the father (they become equal in the limit $u \rightarrow 0$). The network has no idea that a father may exist: it has seen only one of the sons and cannot draw any conclusion about his relatives.

Following the presentation of other $M-1$ sons, the ICC of the firstborn increases monotonically going toward the asymptotic value c_s :

$$\rho_{s1} = (c + (1 - c)q_+\lambda^p)\mu_s^M + c_s(1 - \mu_s^M)$$

The ICC of a generic brother, immediately after his presentation, is greater than the ICC of the eldest brothers at the time of their presentation. In fact when the network had been presented the first sons, there was nothing in the synaptic structure correlated with the common father. The value of the ICC of a son immediately after his presentation increases and tends to the upper bound $c_s(1 - q_+\lambda^p) + q_+\lambda^p$. The ICC of the stimuli which are closer to this maximum decreases immediately after the presentation of the stimuli themselves, and tends to the asymptotic value of all the sons c_s (from above).

In the meantime the ICC of the father is rapidly increasing, in spite of the fact that the father is never presented. In fact the correlation between the father and the sons is greater than that among the brothers. At the beginning the firstborn and some of the eldest brothers prevail, but as the stock grows, the father's ICC can surpass the ICC of all his sons, even that of the youngest sons. The ICC of the father behaves as

$$\rho_f = c\mu_f^P + c_f(1 - \mu_f^P)$$

where P is the total number of sons presented to the network.

The ICC of the eldest 7 sons and of the father is plotted against the number of presented sons in figure 1(A). At the beginning the ICC of the sons prevails (the firstborn dominates), but after the presentation of 5 descendants, the ICC of the father surpasses the ICC of all the sons and tends very rapidly to the asymptotic value.

Following the presentation of 90 sons (figure 1(B)) the ICC of the father is close to its maximum and almost constant, while the ICC of the firstborn is still approaching the asymptotic value c_s from below. The ICC of youngest sons starts from a value which is lower than the firstborn's ICC but then, immediately after their presentation, they jump to their maximum value and, finally decrease toward c_s (from above).

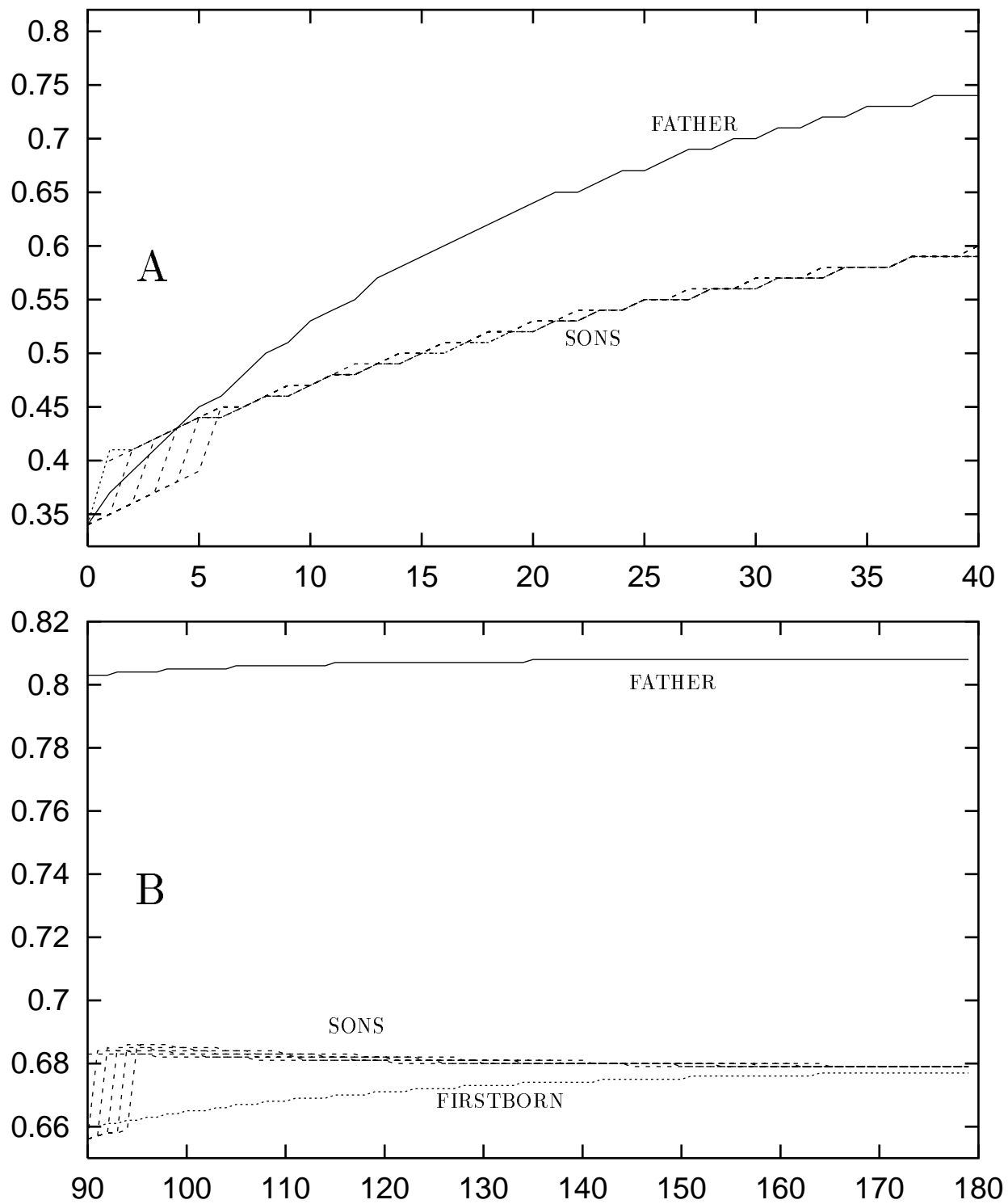


Figure 1: A: ICC of the firstborn (dotted line), the first 6 sons after the firstborn (dashed lines), and of the father (solid line) vs the number of presented sons. B: ICC of firstborn, of the father and of the sons number 90 to 96. See the discussion in the text.

3.5 Condition for a dominating father

The asymptotic value of the father's ICC must be greater than the maximum ICC of any sons. So we have to compare c_f with the ICC of a son immediately after his presentation, following the presentation of an infinite number of eldest brothers. In formal terms:

$$c_f > c_s (1 - q_+ \lambda^p) + q_+ \lambda^p$$

We will consider the case of low coding level ($f \ll 1$). As a consequence we can approximate u with m and U with m^2 . Moreover we assume that $q_- = \mathcal{O}(f)$ (if this condition is not satisfied, the asymptotic distribution cannot preserve any trace of the presented stimuli¹) so that: $\mu_f \simeq (1 - q_+ m^2) \lambda^p$, $\mu_s \simeq (1 - q_+ m^4) \lambda^p$. If we solve the inequality in the variable $r = m^2$, we obtain:

$$-r^2 + (1 - q_+ \lambda^p)r - (1 - \lambda^p)^2 > 0 \quad (4)$$

This implies that the overlap m must lie in the interval $[m_{min}, m_{max}]$ where

$$m_{min,max}^2(q_+, p) = \frac{1}{2} \left(1 - q_+ \lambda^p \pm \sqrt{(1 - q_+ \lambda^p)^2 - 4(1 - \lambda^p)^2} \right)$$

In fact the son can prevail in two cases:

1. $m < m_{min}(q_+, p)$: the brothers are practically uncorrelated. As a consequence they cannot have a common father, there is no family. So the youngest son is always the winner because of his age. We shall label this region by **US** (Uncorrelated Sons)
2. $m > m_{max}(q_+, p)$: the sons are almost identical to the father. The difference between m and m^2 is negligible and the ICC of the youngest brother is always the greatest because it is the last stimulus imposed to the network. We shall refer to this region as to **SCS** (strongly correlated sons).

The upper and the lower bound of m depend on two parameters: p and q_+ . In order to have two real solutions for the equation 4, we must impose the following condition:

$$-(4 - q_+^2) \lambda^{2p} + 2(4 - q_+) \lambda^p - 3 > 0$$

This condition limits the space of variation of p and q_+ . If we solve the inequality in λ^p , we get the condition:

$$\lambda^p > \Lambda_{min} = \frac{4 - q_+}{4 - q_+^2} - \frac{\sqrt{4q_+^2 - 2q_+ + 4}}{4 - q_+}$$

(the upper bound of λ^p is never violated as $\Lambda_{max} > 1$). This constraint limits the number p of patterns which can be interposed between two sons:

$$p < \frac{\log \Lambda_{min}}{-\log(1 - q_+ f^2)}$$

Also this fact is rather intuitive: it says that if the number of uncorrelated patterns between two sons is too high, it is pointless to show a large number of sons in order to extract the structure of the father because the network loses all the information about the eldest sons. In fact when a new son is presented to the network, the synaptic structure has already forgotten everything about the eldest brothers.

m_{min} and m_{max} are plotted in figure 2 as a function of q_+ , assuming for λ^p an intermediate value: $\lambda^p = (1 + \Lambda_{min})/2$. Each value of λ and of q_+ fixes the number of patterns p that can be interposed between two sons. This number decreases as q_+ tends to 1.

The same curve is plotted in figure 3 when p is fixed ($p = 10, 100, 1000$). There is a critical value of q_+ above which there is no set of parameters' values allowing for a dominating father (λ^p exits from the permitted interval). For q_+ greater than this critical value we choose $\lambda^p = \Lambda_{min}$. In this case the strip reduces to a line ($m_{min} = m_{max}$) and the maximum p is determined by the expression $p = \log(\Lambda_{min}) / \log(1 - q_+ f^2)$ (the dotted line represents p , normalized in such a way that 1 corresponds to (from top to bottom): $p = 10, 100, 1000$). In figure 4, for q_+ greater than its critical value, we reduce p according to $\lambda^p = (\Lambda_{min} + 1)/2$. In this case p is smaller but the strip does not collapse in a line.

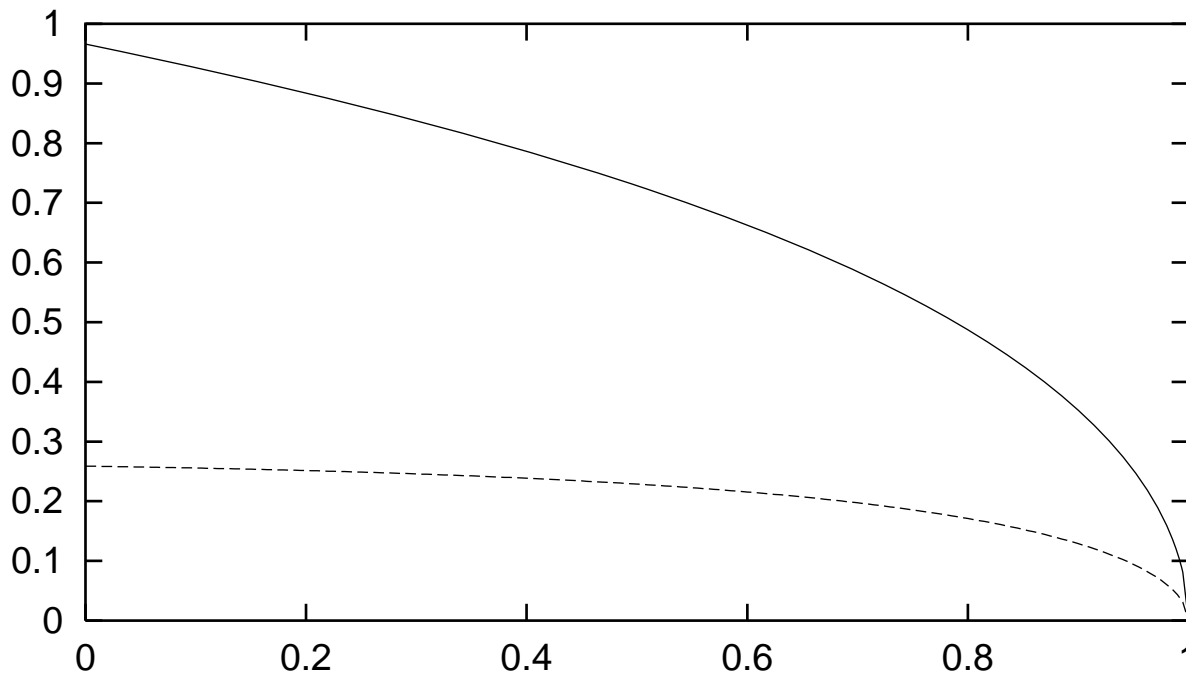


Figure 2: m_{min} (dotted line) and m_{max} (solid line) as a function of q_+ with $\lambda^p = (1 + \Lambda_{min})/2$. Inside the strip the father dominates over all the descendants. Note that as $q_+ \rightarrow 1$ the strips becomes narrower and narrower, until it reduces to a point.

4 DISCUSSION

The set of inequalities discussed in the previous section limits the range of variation of three parameters: m , the similarity among the stimuli belonging to the same group (family); q_+ , which is related to the learning speed; and p , the number of uncorrelated stimuli interposable between two successive presentations of different sons. It can be shown¹² that if q_+ is small and $q_- = \mathcal{O}(f)$, then the results we obtained can be extended to the case in which there is correlation among the n different groups of p uncorrelated random patterns. As a consequence p can be regarded as the number of different classes (families) which can be stored in the synaptic structure.

One of the main general conclusions we can draw is that q_+ cannot be 1, otherwise prototype extraction cannot take place, whatever the neural dynamics is. This implies that for an attractor neural network whose synapses have two stable states, either the up and the down transitions must be stochastic.

If the network dynamics is specified, sufficient conditions for prototype extraction can be found following the same procedure.¹²

5 ACKNOWLEDGMENTS

I'm indebted to Peter Kruger for a careful reading of the manuscript and to Nicolas Brunel and Francesco Battaglia for many useful discussions and comments.

6 REFERENCES

- [1] Amit D.J., Fusi S., 1994 Learning in neural networks with material synapses, *Neural Computation*, **6**, 957–982
- [2] Battaglia F.P., Fusi S., 1994 Learning in neural networks with partially structured synaptic transitions, PREPRINT N.1033, University of Rome, Dept. of Physics, submitted to NETWORK

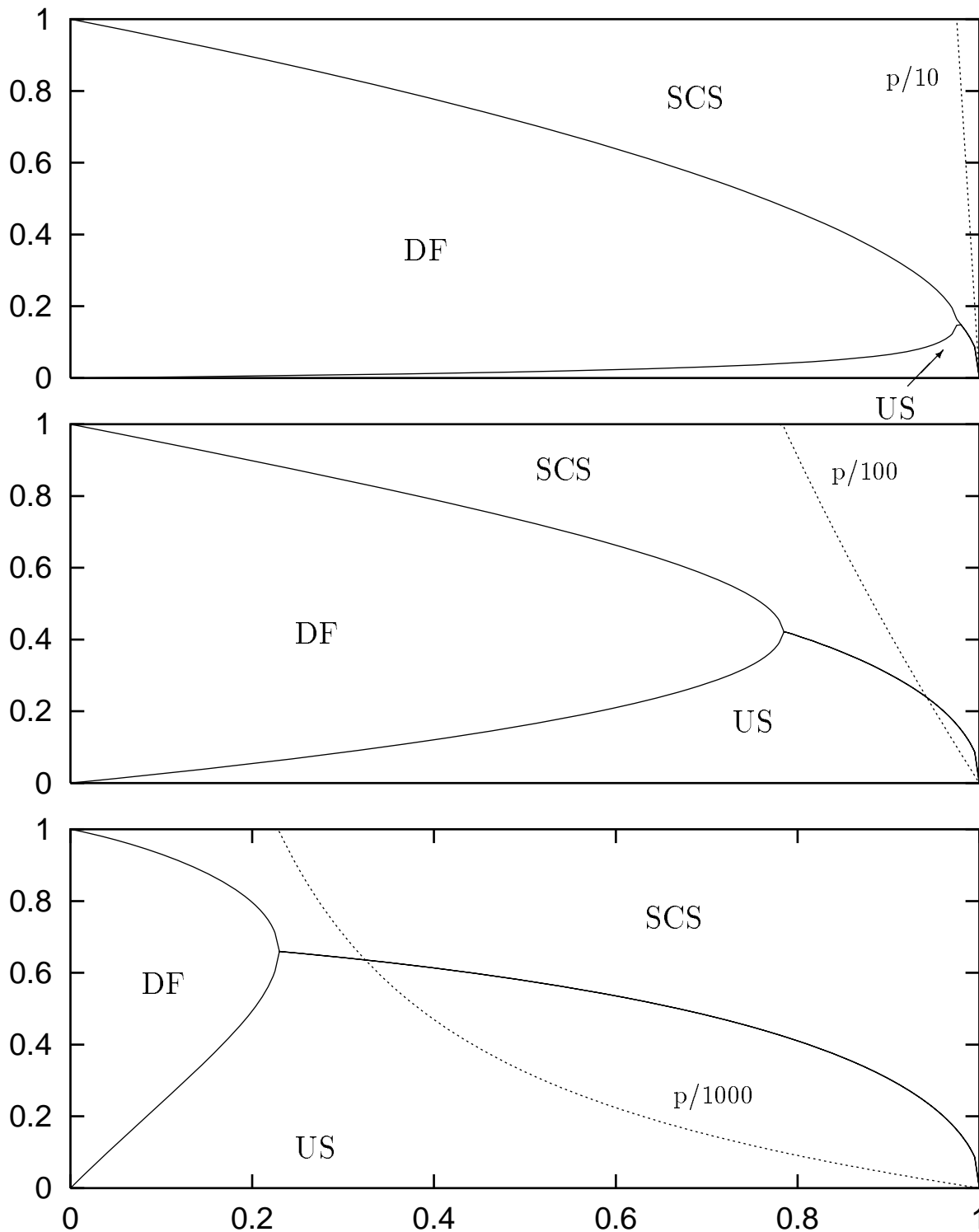


Figure 3: m vs q_+ with fixed p (10,100,1000). $f = 0.05$. As q_+ surpasses its critical value, p is determined by imposing $\lambda^p = \Lambda_{min}$. The strip in which the father dominates (DF) collapses in a line, since $m_{min}(q_+) = m_{max}(q_+)$ and the effective number p of patterns interposed between two sons (dotted lines) decreases as $q_+ \rightarrow 1$. For q_+ below its critical value, the curves $m_{min}(q_+)$ and $m_{max}(q_+)$ part three different regions: SCS (strongly correlated sons), DF (dominating father) and US (uncorrelated sons).

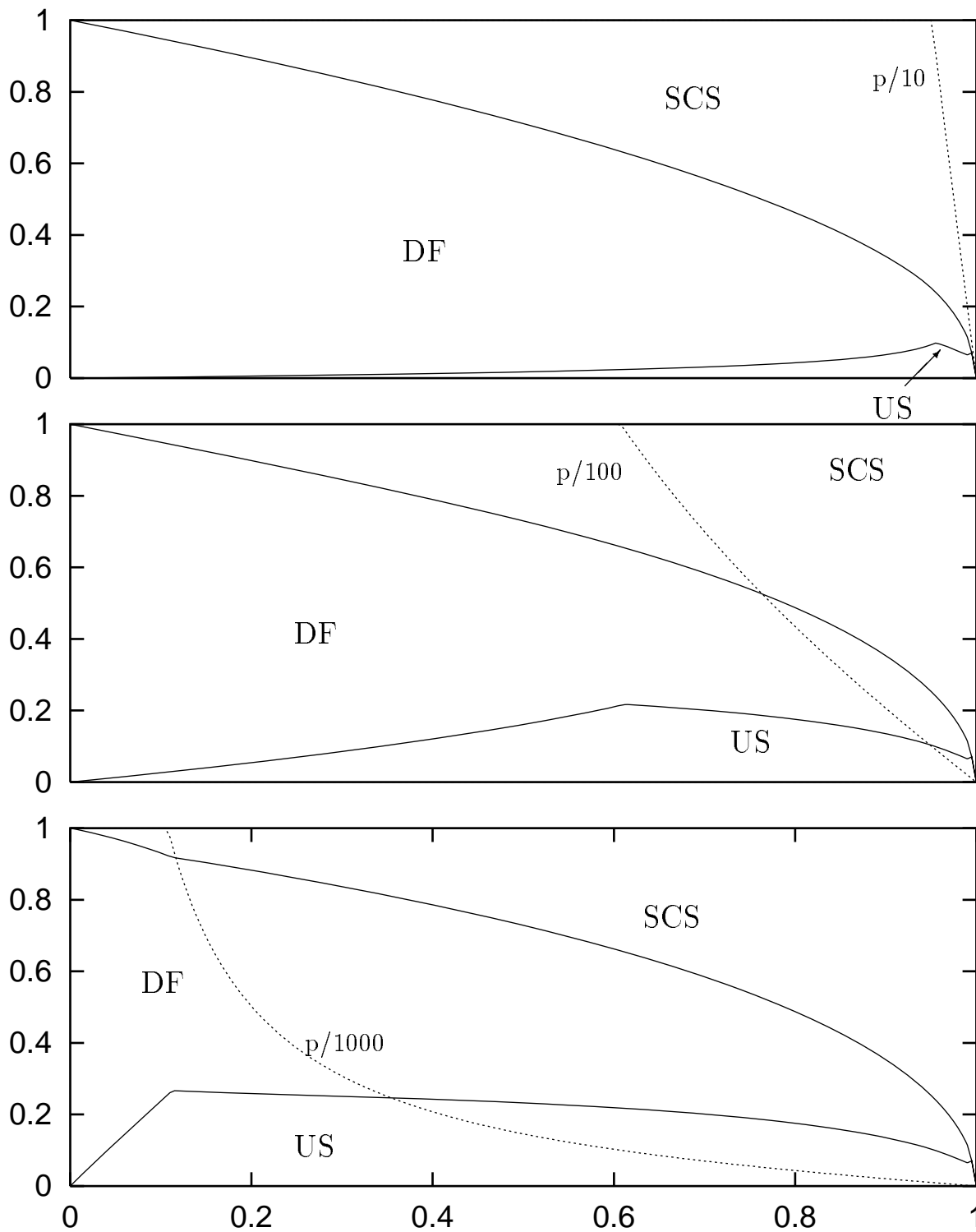


Figure 4: m vs q_+ with fixed p (10,100,1000). $f = 0.05$. For q_+ greater than the critical value, p is given by $\lambda^p = (\lambda_{min}^p + 1)/2$.

- [3] Wong K.Y.M., Kahn P.E., Sherrington D. 1991 A neural network model of working memory exhibiting primacy and recency, *J. Phys A*, **24**, 1119–1135
- [4] Amit D.J., Fusi S., 1992 Constraints on learning in dynamic synapses, *NETWORK* **3**, 443–464
- [5] Nadal J.P., Toulouse G., Changeux J.P., Dehaene S., 1986 Networks of formal neurons and memory palimpsests *Europhysics Letters*, **1**, 532–542
- [6] Parisi G. 1986 A memory which forgets, *J. Phys*, **A19**, L617
- [7] Burgess N., Shapiro J.L., Moore M.A., 1991 Neural network models of list learning, *NETWORK*, **2**, 399–422
- [8] Heskes T.M., Kappen B., 1991 *Phys Rev. A* **44**, 2718
- [9] Cox D.R., Miller H.D. 1965 Theory of stochastic processes (METHUEN & CO LTD London)
- [10] Weisbuch G. Fogelman-Soulié F., 1985 Scaling laws for the attractors of Hopfield model, *J.Physique Lett.*, **2**, 337
- [11] Amit D.J., Brunel N., 1994, Learning internal representations in an attractor neural network with analogue neurons, PREPRINT N.1055, University of Rome, Dept. of Physics, submitted to *NETWORK*
- [12] Brunel N., Fusi S., Kruger P., Storing correlated patterns in attractor neural networks with stochastic dynamic learning, in preparation