Correlated patterns in nonmonotonic graded-response perceptrons

D. Bollé and T. Verbeiren
Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium

(Received 26 July 1999; revised manuscript received 26 January 2000)

The optimal capacity of graded-response perceptrons storing biased and spatially correlated patterns with nonmonotonic input-output relations is studied. It is shown that only the structure of the output patterns is important for the overall performance of the perceptrons.

PACS number(s): 87.10.+e, 02.50.−r, 64.60.Cn

Graded-response perceptrons have been studied intensively in the past years ([1] and references therein). It is found that for nonmonotonic input-output relations interesting retrieval properties are obtained such as an improvement of the optimal capacity (see, e.g., [2]).

The studies mentioned above concern patterns that are chosen to be independent identically distributed random variables with respect to the sites and the patterns. However, in practical applications one has to consider sets of data with internal structure. While the effects of bias and correlations on the optimal capacity have been studied before for monotonic input-output relations ([3–6] and references therein) they have not yet been reported on for nonmonotonic ones. This is the purpose of this Brief Report.

The graded-response perceptron maps a collection of input patterns $\{\xi_i^\mu;1\leq i\leq N\}$, $1\leq\mu\leq p=\alpha N$, with α the capacity, onto a corresponding set of outputs ζ^μ via

$$\zeta^\mu=g(h^\mu),\quad h^\mu=\frac{1}{\sqrt{N}}\sum_j J_j\xi_j^\mu. \tag{1}$$

Here g is the input-output relation of the perceptron. In Eq. (1) h^μ is the local field generated by the inputs. The J_j are the couplings of the perceptron architecture. We focus our attention on general input patterns specified by

$$\langle\xi_i^\mu\rangle=m,\quad \langle\xi_i^\mu\xi_j^\nu\rangle=\delta_{\mu\nu}C_{ij}. \tag{2}$$

The matrix C formed by the elements C_{ij} is taken to be symmetric and positive. We specifically consider correlations with $m=0$ and general C and correlations with $m\neq 0$ and $C_{ij}=\delta_{ij}(v+m^2)$. The latter are called biased patterns and v is the variance of the input distribution.

In order to compute the available Gardner volume [3] in J space we consider the following condition on the local fields:

$$h^\mu\in I^\mu\equiv\{x;g(x)=\zeta^\mu\}, \tag{3}$$

where in general

$$I^\mu=\cup_{j=1}^{r^\mu}I_j^\mu=\cup_{j=1}^{r^\mu}[l_j^\mu,u_j^\mu] \tag{4}$$

form a collection of intervals, not necessarily simply connected, with l_j^μ, u_j^μ the lower and upper bounds of the jth subinterval and r^μ the number of subintervals defined by the pattern ζ^μ. We remark that for monotonic input-output relations, $r^\mu=1$. Following the standard Gardner analysis [3] we use the replica technique to calculate $v=\lim_{N\to\infty}N^{-1}\langle\langle\ln V\rangle\rangle$ with V the fractional volume in J space with spherical normalization and $\langle\langle\ \rangle\rangle$ the average over the statistics of inputs and outputs.

The order parameters occuring in this calculation for correlated patterns with $m=0$ are [4,5]

$$q_{\lambda\lambda'}=\frac{1}{N}\sum_{j,j'}C_{jj'}J_j^\lambda J_{j'}^{\lambda'},\quad \lambda<\lambda', \tag{5}$$

$$Q_\lambda=\frac{1}{N}\sum_{j,j'}C_{jj'}J_j^\lambda J_{j'}^\lambda,\quad \lambda,\lambda'=1,\ldots,n \tag{6}$$

with n the number of replicas. Since the set of general fixed-point equations leading to the optimal capacity α_c (obtained when $V\to 0$) in the replica-symmetric (RS) approximation has been discussed already in [4,5] (for a simple perceptron with the sign function as an input-output relation) we do not write out explicitly the analogous formula for the graded-response perceptron with correlated input and binary output but nonmonotonic input-output relations. For technical reasons the latter are taken to be odd in the field. We just mention that the essential difference is a splitting of the integrations in regions of the form $[(u_{j-1}+l_j)/2,l_j]$ and $[u_j,(u_j+l_{j+1})/2]$, corresponding to the collection of intervals (4) [compare Eq. (8) in [2]]. No closed form for α_c is possible and the solution of these fixed-point equations is rather tedious.

For biased patterns [$m\neq 0$ and $C_{ij}=\delta_{ij}(v+m^2)$] the order parameters read

$$q_{\lambda\lambda'}=\frac{1}{N}\sum_j J_j^\lambda J_j^{\lambda'},\quad \lambda<\lambda', \tag{7}$$

$$M_\lambda=\frac{1}{\sqrt{N}}\sum_j J_j^\lambda,\quad \lambda,\lambda'=1,\ldots,n. \tag{8}$$

Since in this case a closed form for α_c is possible and its structure is interesting for analyzing the effects of the bias m, we write it down explicitly in a first-step replica-symmetry-breaking approximation (RSB1). Applying the Parisi scheme [2,9] we find

$$\alpha_c^{RSB1} = \min_{P,q_0} \max_{M} \frac{-\ln[1+P(1-q_0)] - \dfrac{Pq_0}{1+P(1-q_0)}}{2\left\langle \int Dz_0 \ln\Psi(I^\mu,q_0,P,z_0)\right\rangle_{\zeta^\mu}} \tag{9}$$

with, dropping the index μ in the sequel,

$$\Psi(I^\mu,q_0,P,z_0) = \sum_{j=1}^{r} \mathcal{L}\left(-\mathcal{B}(l_j), -\frac{1}{2}[\mathcal{B}(u_{j-1})+\mathcal{B}(l_j)], y(l_j)\right) + \mathcal{L}\left(-\frac{1}{2}[\mathcal{B}(u_j)+\mathcal{B}(l_{j+1})], -\mathcal{B}(u_j), y(u_j)\right) + \mathcal{L}(-\mathcal{B}(u_j), -\mathcal{B}(l_j), 0), \tag{10}$$

where $u_0=-\infty$, $l_{r+1}=+\infty$, and the other l_j and u_j depend explicitly on the input-output relation g [recall Eqs. (3) and (4)]. Furthermore,

$$\mathcal{L}(a,b,c) = \int_a^b Dz_1 \exp\left(-\frac{1}{2}Pc^2\right), \tag{11}$$

$$\mathcal{B}(x) = \frac{\mathcal{A}(x)+z_0\sqrt{q_0}}{\sqrt{1-q_0}}, \quad \mathcal{A}(x) = \frac{x-mM}{\sqrt{v}}, \tag{12}$$

$$y(x) = \mathcal{A}(x) + z_0\sqrt{q_0} + z_1\sqrt{1-q_0}, \tag{13}$$

with $Dz = dz(2\pi)^{-1/2}\exp(-z^2/2)$ the Gaussian measure. For zero bias we find back the results given in Eq. [2].

Two input-output relations are studied for comparison. The piecewise linear one

$$g_L(x) = \begin{cases} \mathrm{sgn}(x), & |x| \geq 1/\gamma_L \\ \gamma_L x, & |x| < 1/\gamma_L \end{cases} \tag{14}$$

is a prototype of a general monotonic function, and the reversed wedge [7,8]

$$g_{\mathrm{RW}}(x) = \mathrm{sgn}[(x+1/\gamma_{\mathrm{RW}})x(x-1/\gamma_{\mathrm{RW}})] \tag{15}$$

is an example of a nonmonotonic one. Here $\gamma_L, \gamma_{\mathrm{RW}}$ are called gain parameters. In the sequel we use γ without subscript when the results are valid for both types of input-output relations.

Each of the above functions is an example of a class of input-output relations that can be parametrized in the following way. Consider a certain class and take two functions g and $\tilde{g}$ then $(1/\gamma)I^\mu = (1/\tilde{\gamma})\tilde{I}^\mu$. Other examples satisfying this relation include, e.g., the *tanh* function, the inverse linear function, and the sawtooth function, each with the slope as a relevant parameter.

Biased patterns. First, we study biased input and output patterns chosing as their probability distribution

$$\rho(x) = \frac{1+m}{2}\delta(1-x) + \frac{1-m}{2}\delta(1+x), \tag{16}$$

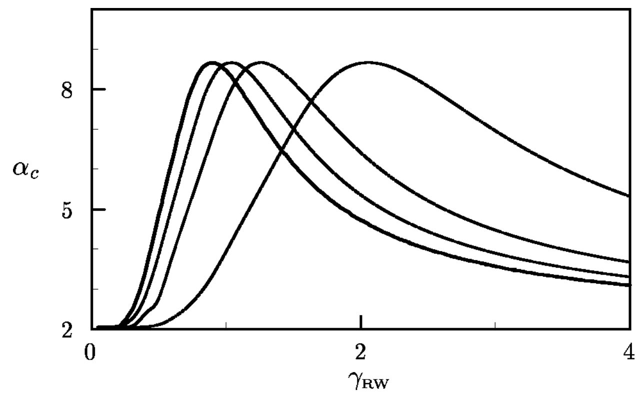

FIG. 1. The optimal capacity α_c of g_{RW} as a function of γ_{RW} with $m_o=0.2$ and, from right to left, $m_i=0.9,0.7,0.5$ and the scaled result [Eqs. (17) and (18)] for these curves.

where m can be different for input and output, thus defining m_i and m_o. Without loss of generality we take the bias parameters m to be positive.

We start with some general properties of the perceptrons defined by Eqs. (14) and (15). Comparing the results (9) with those of [2,10], we see that in order to obtain the expressions for biased patterns it is sufficient to substitute the local field h by $(h-m_iM)/\sqrt{v}$, and to perform an extra maximization over M in the expressions for patterns without bias. This tells us that the order parameter M, which indicates the bias in the couplings, as seen in its definition (8), shifts the local field such that condition (3) is optimally satisfied and hence the capacity increases. Furthermore, it naturally introduces two cases: $m_iM=0$ and $m_iM\neq 0$. Whenever m_i or m_o are zero $m_iM=0$. However, $m_iM=0$ does not necessarily imply that m_i or m_o are zero, as we will see explicitly in the case of the nonmonotonic g_{RW}. These points where $m_iM=0$ occur rather exceptionally, though.

A closer inspection of the results (9)–(13) shows that the graded-response perceptron satisfies the following analytic scaling behavior:

$$\alpha_c(m_i,\gamma;m_o) \equiv \alpha_c(f(m_i,\gamma);m_o), \tag{17}$$

$$f(m_i,\gamma) = (1-m_i^2)^{-1/2}\gamma. \tag{18}$$

These results are valid for both monotonic and nonmonotonic input-output relations (see Fig. 1) discussed above. Other input-output relations need a different parametrization in order to give the simple factor γ in Eq. (18). The new insight is that the output statistics is the important quantity determining the performance of the perceptron. In general, increasing the bias in the output results in a nondecreasing optimal capacity.

Concerning the RS stability assured by a negative sign of the replicon eigenvalue λ_R [9] we find that for monotonic nondecreasing input-output relations and unbiased output patterns the following identity holds for $m_o=0$:

$$\mathrm{sgn}[-\lambda_R(m_i,\gamma_L;0)] = \mathrm{sgn}\left[\frac{\partial}{\partial\gamma_L}\alpha_c(m_i,\gamma_L;0)\right]. \tag{19}$$

This relation tells us that varying the input bias does not change the breaking behavior. The latter can also be seen

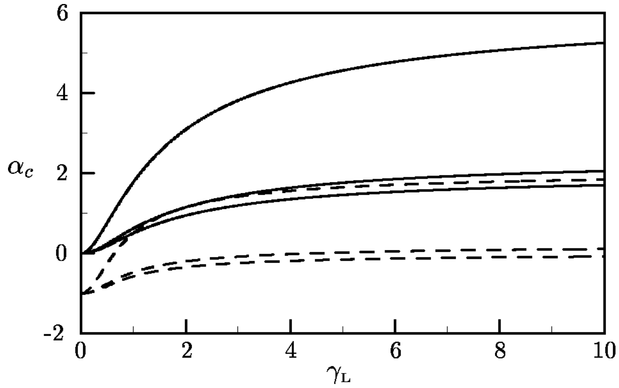

FIG. 2. The optimal capacity α_c (full curves) and the replicon eigenvalue λ_R (dashed curves) for g_L as a function of γ_L with $m_i=0.2$ and, from bottom to top, $m_o=0,0.5,0.9$.

from relation (19) for $m_i=0$ (as given in [10]) together with the scaling relation (17). For nonmonotonic g we know that replica-symmetry is unstable [11].

For the graded-reponse perceptron g_L and pattern distribution (16) we find the following additional results concerning the output bias. For $M=0$ the solution is stable for all values of the bias in input and output, for each γ_L. For $M \neq 0$, from a certain value of m_o onwards the RS solution becomes unstable for a growing interval of γ_L values. This is shown in Fig. 2. For these perceptrons it is known [2,10] that the effect of breaking is small. Although it grows with increasing output bias, it is seen that for $m_o<0.9$, the difference in capacity does not exceed 10^{-2}. We remark that the maximum capacity is reached for $\gamma_L \to \infty$ in agreement with [3].

For the nonmonotonic input-output relation $g_{\rm RW}$ the maximal α_c is obtained for a finite value of $\gamma_{\rm RW}$, as shown in Fig. 3, implying that there exists an optimal choice for the width of the plateaus. This choice depends on the specific parameters of the pattern distribution.

Compared with the monotonic case the overall difference between the RS and RSB1 solution is much bigger. The optimal capacity for the nonmonotonic perceptron is always greater than that for the monotonic one. We note that for $\gamma_{\rm RW}\to 0$ and $\gamma_{\rm RW}\to\infty$ the optimal capacity of $g_{\rm RW}$ approaches that of the sign function. Technically, for every

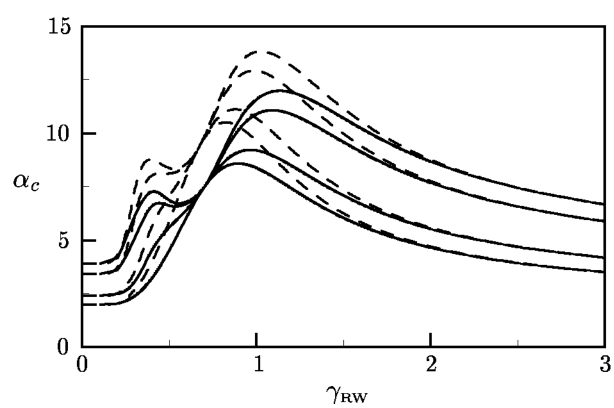

FIG. 3. The RS (dashed curve) and RSB (full curve) optimal capacity α_c of $g_{\rm RW}$ as a function of $\gamma_{\rm RW}$ with $m_i=0.1$ and, from bottom to top, $m_o=0,0.5,0.75,0.8$.

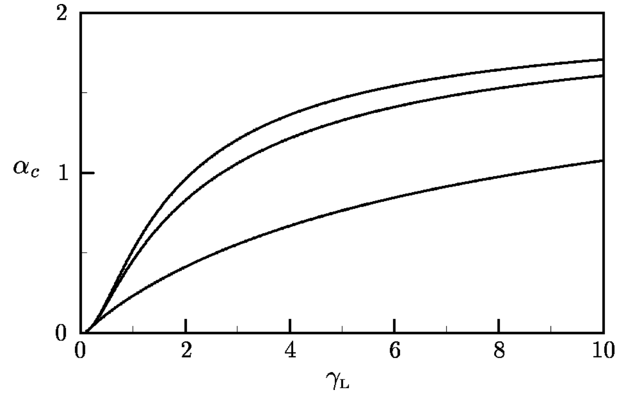

FIG. 4. The optimal capacity α_c as a function of γ_L for g_L with correlated inputs. From top to bottom $S=0,0.5,0.9$.

$m_o>0$ we find that there exist multiple solutions of the relevant fixed-point equations for small values of $\gamma_{\rm RW}$. This is due to the nonmonotonicity of the input-output relation. In that case we take the solution giving the greatest optimal capacity α_c.

A somewhat surprising feature of this perceptron is that a second maximum develops both in the RS and RSB1 solution as a function of $\gamma_{\rm RW}$ for big values of m_o (Fig. 3). Qualitatively the overall behavior of the input-output relation remains the same within RSB1. This is so for all values of the model parameters we have considered but may be a property of the binary output distribution (compare [2] for a uniform output). The difference between RS and RSB1 grows with increasing bias.

Between the two maxima, there is a point where α_c does not depend on the output m_o. This feature is present both in the RS and the RSB1 approximation although for a slightly different value of $\gamma_{\rm RW}$ in RSB1. The underlying reason for this is that the solution of M at these points is zero and that the input-output relation is odd in the local field, such that the output statistics does not influence the optimal capacity of the system. Since changing m_i can be expressed as rescaling $\gamma_{\rm RW}$ in the sense of Eq. (17), the capacity at these points is the same for every value of m_i and m_o. Changing the input distribution (16) by varying the place of the delta peaks in the interval [0,1] shows a similar scaling behavior.

Correlated patterns. Next, following [4], we study correlations in the input patterns that are positive and fall off with the distance between the sites

$$C_{ij}=\exp(-|i-j|/L)\equiv S^{|i-j|}, \tag{20}$$

with L a typical length size. The parameter S is the correlation strength inside one input pattern and varies between 0, corresponding to independent sites and 1 meaning that all spins in a pattern are equal. The spatial structure introduced above induces interesting correlations between the couplings [4].

For the sign function it has been shown [4] that the optimal capacity α_c remains 2, regardless of the inner structure of the inputs. First, for the case of g_L we analyze the critical capacity α_c as a function of γ_L for different correlation strengths S. The results are shown in Fig. 4. By increasing γ_L or decreasing S, α_c increases. In the limit $\gamma_L\to\infty$, g_L be-

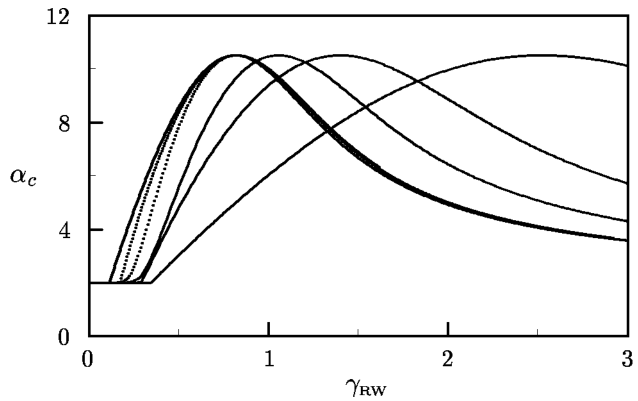

FIG. 5. The optimal capacity α_c as a function of γ_{RW} for g_{RW} with correlated inputs. From right to left $S=0.9,0.7,0.5$ and the scaled results (21)–(22) for these curves.

comes the sign function such that α_c always approaches 2 because of the argument above.

Second, for the nonmonotonic input-output relation g_{RW} the corresponding results are presented in Fig. 5. Several remarks are in order. Technically, the solutions of the relevant saddle-point equations are unique for small values of S for all γ_{RW} but from $S>S_c=0.55$ onwards there exist, as in the case of biased output patterns, multiple solutions in a growing interval in γ_{RW}. Taking again the solution giving the greatest optimal capacity we find the constant part $\alpha_c=2$ (for small γ_{RW} and $S>S_c=0.55$) of the curves in Fig. 5. It seems that for these values of γ_{RW}, the perceptron is not able to benefit from the nonmonotonicity of g_{RW} due to the fact that the order parameter Q remains small.

Looking at Fig. 5 we see that the maximal value of α_c is the same, independent of the correlation strength. This can also be shown using the structure of the fixed-point equations (9)–(13). We find

$$\alpha_c(S,\gamma_{RW})\equiv\alpha_c(f(S,\gamma_{RW})), \qquad (21)$$

where the precise scaling function f can be determined only numerically due to the complex structure of the fixed-point equations. However, we can show analytically that for $\gamma_{RW}\equiv\gamma_c$ defined such that α_c is maximal

$$f(S,\gamma_c)=\left(\frac{1+S^2}{1-S^2}\right)^{1/2}\gamma_c. \qquad (22)$$

These results imply that for the nonmonotonic g_{RW} the inner structure of the patterns does not play an important role in the overall performance of the perceptron.

In this Brief Report we have studied the optimal capacity of a class of graded-response perceptrons storing biased and spatially correlated patterns with nonmonotonic input-output relations using a first-step replica-symmetry-breaking analysis.

The most important results are that a change in the optimal capacity due to bias or correlations in the input can be removed by an appropriate scaling of the parameters defining the graded-response perceptrons. The statistics of the outputs determines the performance of the latter.

The authors would like to thank Dr. G. Vancraeynest for informative discussions and the Fund for Scientific Research Flanders (Belgium) for financial support.

[1] D. Bollé and R. Erichsen, Jr., Phys. Rev. E **59**, 3386 (1999).
[2] D. Bollé and R. Erichsen, Jr., J. Phys. A **29**, 2299 (1996).
[3] E. Gardner, Europhys. Lett. **4**, 481 (1987); J. Phys. A **21**, 257 (1988).
[4] R. Monasson, J. Phys. A **25**, 3701 (1992).
[5] W. Tarkowski and M. Lewenstein, J. Phys. A **26**, 2453 (1993).
[6] J. O. Winkel, B. Schottky, U. Krey, and F. Gerl, Z. Phys. B: Condens. Matter **101**, 305 (1996).
[7] T. L. H. Watkin and A. Rau, Phys. Rev. A **45**, 4102 (1992).
[8] G. Boffetta, R. Monasson, and R. Zecchina, J. Phys. A **26**, L507 (1993).
[9] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[10] D. Bollé, R. Kuhn, and J. van Mourik, J. Phys. A **26**, 3149 (1992).
[11] M. Bouten, J. Phys. A **27**, 6021 (1994).