EUROPHYSICS LETTERS May 1998

Europhys. Lett., (), pp. (1998)

Numerical simulation of a binary communication channel: Comparison between a replica calculation and an exact solution

D.R.C. Dominguez[1](*), M. Maravall[1](**), A. Turiel[1], J.C. Ciria[1](***) and N. Parga[1]

[1] *Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

(received May 1998; accepted -)

PACS. 87.10.+e – General, theoretical, and mathematical biophysics.
PACS. 64.60.Cn – Order-disorder transformations; statistical mechanics of model systems.

Abstract. – The mutual information of a single-layer perceptron with N Gaussian inputs and P deterministic binary outputs is studied by numerical simulations. The relevant parameters of the problem are the ratio between the number of output and input units, $\alpha = P/N$, and those describing the two-point correlations between inputs. The main motivation of this work refers to the comparison between the replica computation of the mutual information and an analytical solution valid up to $\alpha \sim O(1)$. The most relevant results are: (1) the simulation supports the validity of the analytical prediction, and (2) it also verifies a previously proposed conjecture that the replica solution interpolates well between large and small values of α.

Keywords: Statistical physics, information theory, neural networks, unsupervised learning.

Introduction. –

The extraction of sensory information by the brain from a stream of multi-dimensional data may be understood as a process of optimisation of mutual information (MI) [1] or of redundancy [2]. The MI measures the statistical dependence between two random variables [3]. In our case, they correspond to an N-dimensional input signal $\vec{\xi}$ provided by the sensory receptors, and a P-dimensional output $\vec{v}$. More precisely, the MI indicates the amount of knowledge about $\vec{\xi}$ that can be extracted from $\vec{v}$ (See e. g. ref.[3]). With respective probability densities p_ξ and p_v, this is given by

(*) Present address: Instituut voor Theoretische Fysica, K.U. Leuven, 3001, Leuven, Belgium.
(**) Present address: Institute for Theoretical Physics, State University of New York at Stony Brook, NY 11794-3840, USA.
(***) Permanent address: Departamento de Física Teórica. Universidad de Zaragoza.C/ Pedro Cerbuna, 12 50010 Zaragoza. SPAIN

Typeset using EURO-TEX

$$I[p_v, p_\xi] = < \log \frac{p_{v,\xi}(\vec{v}, \vec{\xi})}{p_v(\vec{v}) p_\xi(\vec{\xi})} >_{v,\xi}, \tag{1}$$

here $\log(x) \equiv \frac{\ln(x)}{\ln 2}$. The average $< ... >_{v,\xi}$ is over the joint probability distribution $p_{v,\xi}(\vec{v}, \vec{\xi})$. For instance, if $\vec{v}$ and $\vec{\xi}$ are independent, we have $p_{v,\xi} = p_v \cdot p_\xi$ and so $I = 0$.

The problem of learning the statistical properties of a set of N-dimensional correlated Gaussian inputs is well understood for a linear channel, even in the presence of noise (which is actually necessary to regularise MI)[4]-[6]. A non-linear continuous channel has also been studied in the low-noise limit for rather general transfer functions [7]. It was shown that maximisation of the MI leads to a factorial code. Threshold-linear networks [8, 9] (treated with the replica technique) have also been considered. On the other hand, the binary channel, where the outputs take discrete values (say $v_\mu = \pm 1$), is not well understood, although the problem has been studied using replica-symmetric (RS) statistical-mechanical techniques [10]-[12]. Most interestingly, an analytical solution has been found, and the existence of a large order phase transition as the number of output neurons increases has been suggested [12]. The relevant parameter to describe this transition's occurrence is the ratio between the number of output and of input units, $\alpha = P/N$. The analytical solution holds up to a value of α of order one, beyond which it is not longer correct. On the contrary, the RS solution does not exhibit any transition and gives good approximations at both the small and large α regimes. It has been proved that below some $\alpha \sim 1$ the analytical and the RS solutions are very close. In fact, an expansion in powers of α shows that the two solutions are identical up to $O(\alpha^2)$. In spite of the fact that from the third order the corresponding expansions differ, the numerical agreement up to $\alpha \sim 1$ is excellent (a relative difference of less than 0.9% up to $\alpha = 0.1$). This is due to intriguing cancellations between higher orders.

Here we present simulations with the aim of providing numerical evidence on the validity of the analytical solution. In addition, since the order of the transition is large and the RS solution seems to interpolate well between the small α and the asymptotic regimes, we also compare numerical simulations at several values of α with the replica theory prediction [10].

The Binary Channel. –

We consider a single-layer perceptron, or *channel*, with N continuous input neurons whose states ξ_i define a vector $\vec{\xi} = \{\xi_i\}_{i=1}^N$ representing the *signal* received from the environment. The output layer has P binary neurons the values $v_\mu = \pm 1$ of which compose the vector $\vec{v} = \{v_\mu\}_{\mu=1}^P$, that represents the *code*. Between the signal and the code there is an encoder, given by a set of synaptic couplings $\mathcal{J} \equiv \{\mathcal{J}_{i\mu}\}$.

The inputs take values drawn from an N-dimensional Gaussian probability distribution, unbiased ($< \xi_j >= 0$) and with correlation matrix $\mathbf{C}_X^{ij} =< \xi_i \xi_j >$,

$$p_\xi(\vec{\xi}) = e^{-\vec{\xi} \cdot \mathbf{C}_X^{-1} \cdot \vec{\xi}/2} / \sqrt{2\pi C_X},\ C_X = |\mathbf{C}_X|, \tag{2}$$

which using a convenient shorthand can be expressed as $\vec{\xi} \doteq N(\vec{0}, \mathbf{C}_X)$. (Here X is a correlation parameter between input neurons, to be defined more precisely later). The transfer function is deterministic, so that $v_\mu = \mathrm{sign}(h_\mu)$, where

$$h_\mu = \sum_{i=1}^N J_{i\mu} \xi_i \equiv \vec{J}_\mu \cdot \vec{\xi} \tag{3}$$

and the $\vec{J}_\mu$'s denote the synaptic weight vectors linking the signal $\vec{\xi}$ to each output neuron μ. They form a set of independent random vectors $\{\vec{J}_\mu\}_{\mu=1}^P$, each distributed according to an N-dimensional Gaussian probability, with mean $< J_{i\mu} >= 0$ and correlation matrix $\mathbf{\Gamma}_\mu$, whose elements are $\Gamma_\mu^{ij} =< J_{i\mu} J_{j\mu} >$. This means that $\vec{J}_\mu \doteq N(\vec{0}, \mathbf{\Gamma}_\mu)$.

We are mainly interested in computing the averaged mutual information per input unit in the thermodynamic limit, *i.e.* $i \equiv \lim_{N\to\infty} \frac{1}{N} < I[p_v, p_\xi] >_{\mathcal{J}}$. A useful result to have in mind is that the MI in eq.(1) is the difference

$$I[p_v, p_\xi]_{\mathcal{J}} = H[p_v] - H[p_{v|\xi}], \tag{4}$$

where $H[p_v] = - < \log p_v >_v$ is the entropy of the output, while $H[p_{v|\xi}] = - << \log p_{v|\xi} >_{v|\xi} >_\xi$ is the conditional entropy of the output given the input (averaged over the input). The code $\vec{v}$, given a fixed signal $\vec{\xi}$, has the conditional probability $p_{v|\xi} = p_{v,\xi}/p_\xi$. Since a deterministic channel clearly has zero conditional entropy, in that case the MI reduces to the output entropy.

Analytical results for an example. –
As an example we consider a Gaussian input distribution with two-point correlation $\mathbf{C}_X$ given by $C_X^{ij} = X^{|i-j|}$. The input neurons are then less (more) spatially correlated if $X \ll 1$ ($X \sim 1$). For $X = 1$, $\mathbf{C}_X$ is a singular matrix, while for $X \to 0$ $\mathbf{C}_X$ tends towards the identity matrix. The correlations between two synapses converging to the same output, Γ_μ^{ij}, are chosen to be equal for all output neurons and normalised to 1, *i.e.* $\Gamma_\mu^{ij} = \Gamma^{ij} = \delta_{ij}$, $\forall\mu$.

For small α, the MI can be expanded as $i(\alpha) \sim a_0 + a_1\alpha + a_2\alpha^2 + a_3\alpha^3$. One trivially obtains that $a_0 = 0$ and $a_1 = 1$. Both the RS solution (see [10]) and the analytical solution (see [12]) give the same value of a_2,

$$a_2(X) = \frac{-\gamma}{\pi^2 \ln(2)} \tag{5}$$

where $\gamma \equiv \frac{1+X^2}{1-X^2}$. On the other hand, the two techniques disagree in their predictions for a_3. The RS method yields

$$a_3^{RS}(X) = \frac{(6 - 12/\pi)\gamma^2 - 2}{3\pi^3 \ln(2)}, \tag{6}$$

while (following the methods of [12]) one can easily verify that the analytical technique gives:

$$a_3^{An}(X) = \frac{6\gamma^2 - 2}{3\pi^3 \ln(2)}. \tag{7}$$

The RS solution also gives rise to predictions for strongly correlated inputs. For $X \sim 1$ one obtains:

$$i \sim K(1-X)^{1/3}\alpha^\nu,\ \nu \equiv 2/3. \tag{8}$$

Finally, the same solution shows logarithmic behaviour for large α:

$$i \sim \log\alpha + \frac{1}{2\ln 2} + \log 0.72 + \frac{1}{2N}\mathrm{Tr}[\log \frac{\mathbf{C_X}}{N}]. \tag{9}$$

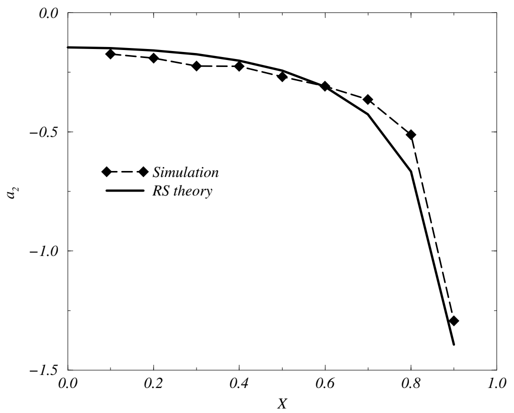

Fig. 1. – The coefficient $a_2(X)$ obtained from simulations compared with the RS prediction.

Method. –
The first step of the simulation entails choosing a coupling sample $\mathcal{J}$ at random. A given choice of $\mathcal{J}$ will be labelled as the sample s, and the total number of samples will be denoted by S. Next, signals $\vec{\xi}$ are drawn in order to obtain several codewords $\vec{v}$. A histogram $p_v^s(\vec{v})$ is then constructed from these states, representing the probability $p_v(\vec{v})$ and allowing us to estimate the output entropy:

$$i_s = -\frac{1}{N} \sum_{\{v_\mu = \pm 1\}} p_v^s \log p_v^s. \tag{10}$$

Clearly, this approximation to the true code entropy $H[p_v]$ will improve as the number of drawn signals increases. In practice, we evaluated $p_v^s(\vec{v})$ using about 100 different input states $\vec{\xi}$. The final step is to calculate the average over an ensemble of $\mathcal{J}$, to get $i = \frac{1}{S}\sum_{s=1}^{S} i_s$.

Results from simulations. –
Here we present studies of three relevant regions in parameter space (α, X): (1) α small ; (2) intermediate values of α and strong correlations, $X \sim 1$; and (3) α large. The results are compared with the theoretical predictions, eqs.(5)-(9). The self-averaging property of the MI is also analysed.

In the **small** α **case**, we calculated $i(\alpha)$ for values of α ranging from 0.05 to 0.2: we fixed $P = 10$ and took a variable number of inputs in the interval $N = 50, ..., 200$. We repeated this process for several values of the correlation parameter $X = 0.1, 0.2, ..., 0.9$. Using that $i(\alpha = 0) = 0$ when $N \to \infty$, we performed a linear regression on i, $i(\alpha, X)/\alpha = a_1(X) + a_2(X)\alpha$, and obtained the coefficients a_1, a_2 as a function of X. For all values of X, $a_1 \sim 1$, which is in agreement with the theory. The function $a_2(X)$ is plotted in fig.1, in comparison with the theoretical prediction (eq.(5)). The good agreement indicates that the thermodynamic limit solution is a good estimation even for N not very large, as long as one is in the low-loading expansion.

In order to evaluate a_3, we set $a_1 \equiv 1$ (to avoid increasing the error through a larger number of parameters to be fitted). We analysed the MI only for the value $X = 0.5$. From eq. 5 the theoretical value $a_2 = -0.244$ can be obtained, and from eqs. 7 and 6 we have $a_3^{An} = 0.227$ and $a_3^{RS} = 0.063$ respectively. The simulation itself was done using several values of P, and for each of them a linear regression was performed using $[i(\alpha)/\alpha - 1]/\alpha = a_2 + a_3\alpha$.

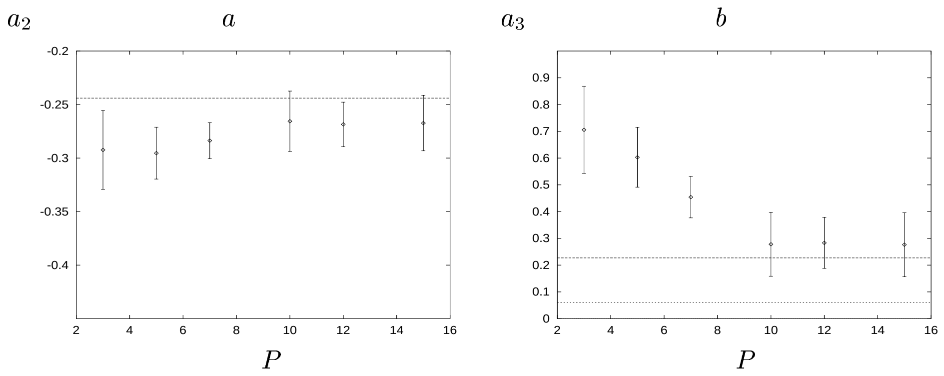

Fig. 2. – The coefficients a_2 (a) and a_3 (b) for several P's, for $X = 0.5$. The straight horizontal lines represent the theoretical values; $a_2^{RSA} = a_2^{An} = -0.244$ and $a_3^{RSA} = 0.06$, $a_3^{An} = 0.23$. The exact solution is in good agreement with the data, while the RSA coefficient lies outside the error bars.

In this case the corrections due to finite size effects are noticeable. To observe convergence to the asymptotic regime we considered several values of P, namely $P = 3, 5, 7, 10, 12$ and 15. We averaged for each over, respectively, $S = 2500, 1000, 200, 100, 50$ and 20 samples of $\mathcal{J}$. The number of samples S was taken larger as the number of outputs P became smaller, to compensate for the lack of statistics. Numerical evaluation of a_3 is extremely costly in computational time; this is the reason why we restricted ourselves to a single value of X and did not consider values of P larger than 15.

The result is that our simulation is in agreement with the analytical prediction. As we see in fig.2b, a_3^P converges to $a_3 \sim 0.28$ as P increases. Given the error bars, this result is compatible with a_3^{An} but *excludes* a_3^{RS}. For the sake of comparison we have included fig. 2a where the same numerical analysis is done for a_2.

The results for **strongly correlated inputs** are shown in fig. 3a. There is good agreement between the simulation and the exponential behaviour of eq.(8): $\nu = 0.662 \sim 2/3$.

The results for the **large α limit** are presented in fig.3b. The logarithmic behaviour predicted in eq.(9) is observed.

To verify that the MI is self-averaging we calculated its mean-square deviation $\Delta(i)$ over the samples and made a fit to the form $\frac{1}{\sqrt{N}}$. The good agreement with this expression can be seen in fig. 3c. This shows that the methods of statistical mechanics are appropriate to studies of the information in binary channels.

Conclusions. –

Our main result refers to the comparison between the RS [10] and analytical [12] solutions. The difference between them can be seen in a small-α expansion. Numerical simulation confirms that the two solutions coincide up to second order. At the next order the solutions are different (see eqs. 6-7), and the simulation excludes the replica calculation while it is in agreement with the analytical one (see fig. 2).

We have also verified the conjecture that the RS solution is a good interpolation between the small and the large α behaviors [12]. In particular, the simulation shows that for intermediate values of α and strongly correlated inputs the MI behaves as $i \sim \alpha^{2/3}$ (see fig.3a). Moreover, in fig.3b we see that the expected logarithmic behaviour for the large α case fits the MI very well.

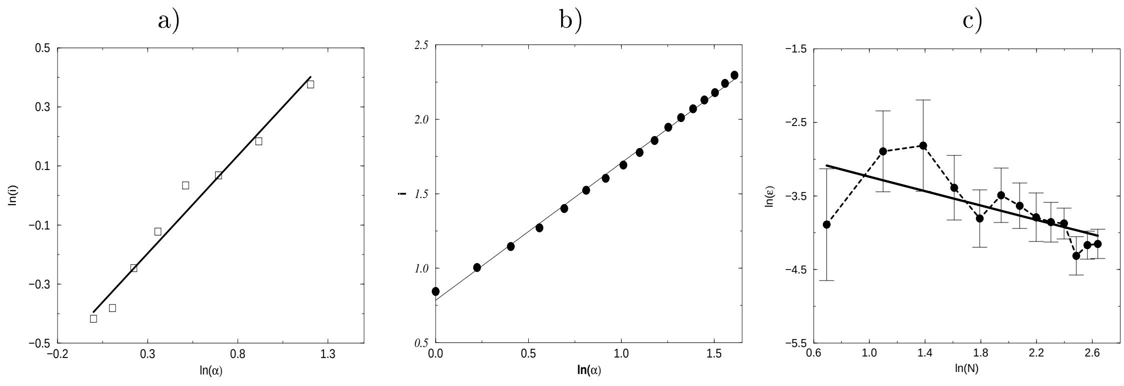

Fig. 3. – **a)** The behaviour of the MI for strongly correlated inputs ($X = 0.9$). $P = 10$ and $N = 3, ..., 10$. The squares are obtained from simulation, while the plotted line is: $\ln(i) = -0.395 + 0.662\ln(\alpha)$. **b)** The MI for large α, with $N = 4, P = 1, ..., 20$ and $X = 0.1$ ($\alpha \leq 5$). The dots are obtained from simulation, while the plotted curve is $i = 0.78 + 0.92\ln(\alpha)$. **c)** The self-averaging property. We took $X = 0.9$ and $P = N = 2, ..., 14$ ($\alpha = 1$). The linear regression gives $\ln[\Delta(i)/i] = -2.5 - 0.49\ln(N)$, where $\Delta^2(i) = \frac{1}{S}\sum_{s=1}^{S}[i_s - i]^2$.

Acknowledgments. – This work was supported by a Spanish grant PB 96-47. Antonio Turiel is financially supported by an FPI grant from the Comunidad Autónoma de Madrid, Spain. M. Maravall was supported by a Beca de Colaboración from the Spanish Ministry of Education. D.R.C. Dominguez thanks the K.U. Leuven for a research fund (grant OT/94/9).

REFERENCES

[1] Linsker R.; *Computer*, **21** (1988) 105-17.
[2] Barlow H.B.; "Possible principles underlying the transformation of sensory messages"; *Sensory Communications* Cambridge, MA: MIT Press, 1961.
[3] Shannon C.E. and Weaver W.; "The mathematical theory of communications" Urbana, IL: University Illinois Press, 1994.
[4] Linsker R.; *Neural Computation* **4** (1992) 691.
[5] Atick J.J. and Redlich A.N.; *Neural Computation* **4** (1992) 196-210.
[6] Campa A., Giudice P., Parga N. and Nadal J.-P.; *Network* **6** (1995) 449.
[7] Nadal J.-P. and Parga N.; *Network* **5** (1994) 565.
[8] Treves A.; *J. Computational Neuroscience* **2** (1995) 259.
[9] Schultz S. and Treves A.; *Phys. Rev E* **57** (1998) 3302-3310.
[10] Nadal J.-P. and Parga N.; *Network* **4** (1993) 295.
[11] Koroutcheva E., Parga N. and Nadal J.-P.; *Network* **8** (1997) 405.
[12] Turiel A.M., Koroutcheva E. and Parga N.; "On the mutual information of a stochastic binary channel: The validity of the Replica Symmetry Ansatz". Submitted to *J of Physics A*.