Computational capability of classical and quantum spin systems

1993 J. Phys. A: Math. Gen. 26 L201

(http://iopscience.iop.org/0305-4470/26/5/005)

J. Phys. A: Math. Gen. 26 (1993) L201–L207. Printed in the UK

LETTER TO THE EDITOR

Computational capability of classical and quantum spin systems

Ido Kanter and Eli Eisenstein

Department of Physics, Bar-Ilan University, Ramat Gan 52100, Israel

Received 13 April 1992, in final form 18 December 1992

Abstract. A quantitative information theory is developed for both classical and quantum spin systems. The theory is examined by heavy numerical simulations, which show that the information embedded in each weight of the system increases with the size of the system. The difference between the notions of capacity and information is examined carefully and gives an explanation to the surprising result that the capacity per bit can be greater than one.

The application of statistical physics methods to the investigation of neural network models has recently attracted much attention [1]. In the simplest model, the system is composed of N binary elements, represented by Ising spin variables, S_i, which could take the values ± 1 and connected by the weights (synapses) $\{J_{ij}\}$. The dynamic of the network in the zero temperature limit is given by $S_i(t+1) = \mathrm{sgn}[\Sigma_{j\neq i} J_{ij} S_j(t)]$. The basic task of these networks is to find $\{J_{ij}\}$ for which a set of given random patterns, $\xi_i^\mu = \pm 1$ with equal probability ($i = 1, \ldots, N$, $\mu = 1, \ldots, p$), are fixed points of the dynamics. In the case that such a set of $\{J_{ij}\}$ exists, the p patterns are said to be embedded in the network. The critical capacity of the network is defined as the maximal number of random patterns, P_c, which can be embedded in the network.

One of the notions which played an important role in the last few years in estimating the computational capability of networks was the maximal number of embedding patterns per weight. This quantity was calculated analytically for various models, where almost all of them were asymmetric networks, $J_{ij} \neq J_{ji}$, with a continuous space for the weights [2]. However, the capacity per weight was recently found to be unbounded [3]. Hence, this notion is not a well defined measure to estimate the capability of networks, independent of their architectures. It was recently found that the right notion to estimate the computational capability of networks is not the capacity per weight but rather the capacity per bit [4]. In the case of discrete weights the computational capability of a network is defined as the number of bits which are embedded in the patterns divided by the number of bits which are necessary for the representation of the weights [4]. In this picture, a network is an 'information engine' which converts information from one type of presentation to another [5]. More precisely, the network translates information from the presentation in the space of the weights to the presentation in the space of the spins.

The capacity of discrete *symmetric* systems was recently calculated analytically, and surprisingly it was found that the capacity per bit can be even greater than one [5]. Therefore, it is clear that the capacity per bit is the right measure for feedforward networks, but is the wrong measure to estimate the embedded information in symmetric systems. In the present situation it is fair to say that the right measure, which is

0305-4470/93/050201+07$07.50

independent of the model, to estimate the computational capability of symmetric systems is in question. This question and the relation between the right measure for symmetric systems and the capacity per bit is at the centre of this work. A deeper understanding of these questions is necessary in order to understand the fundamental differences, advantages and disadvantages, between the computational capability of physical laws over asymmetric laws. Beside this goal, a consistent quantitative information theory is also necessary to calculate the efficiency of nature to convert information from the space of the weights to information in the space of the particles (spins in the following discussion).

In this work a quantitative information theory is built for both classical and quantum spin systems in the zero temperature limit. This theory is examined by heavy numerical simulations. The differences between the maximal embedded information in classical and quantum systems are examined, and a comparison between the notions information and capacity of symmetric systems is discussed.

The classical spin systems which are examined below are governed by the classical Hamiltonian

$$H_c = -\sum_{i \neq j}^{N} J_{ij} S_i S_j \tag{1}$$

where N is the size of the system and the weight J_{ij} and the spin S_i can take any discrete values. For simplicity, only the case of Ising spins, $S_i = \pm 1$ and binary weights, $J_{ij} = \pm 1$, is examined. The quantum spin Hamiltonians which are examined are

$$H_q = \sum_{i \neq j}^{N} J_{ij} \boldsymbol{S}_i \cdot \boldsymbol{S}_j \tag{2}$$

where the quantized values of the operator $\boldsymbol{S}_i^2$ can be chosen arbitrarily. Nevertheless, for simplicity, we shall restrict ourselves to the spin-$\frac{1}{2}$ case with binary weights, $J_{ij} = \pm 1$. In the following, first the classical case equation (1), is discussed, then the quantum case is examined and compared to the classical one.

The only time independent embedded information of classical systems is the fixed points of the dynamics, states which are stable to a single spin flip. In the case of a fully connected classical system of size N, the total number of weights is $N(N-1)/2$ and hence there are $N_c = 2^{N(N-1)/2}$ different configurations (in the binary case). A configuration embedded a *new information* if, and only if, its time independent information is not included in the information of any of the other configurations. More precisely, for each configuration one can assign a symbol which stands for a set (list) of the stable patterns for this particular configuration [6]. Each configuration has at least one fixed point and neglecting the effect of a global inversion symmetry of the Hamiltonian, equation (1), the number of fixed points is bounded from above by 2^{N-1}. A configuration A, for instance, contains new information if, and only if, the set of patterns represented by the A's symbol is not a subset or equal to any other symbol B, where $A \neq B$. The computational ability of the system is defined by the minimal number of independent symbols, N_s. Each one of these symbols contains new information and the information of all the other symbols is included in the N_s symbols. This number can be represented by $\log_2 N_s$ bits. On the other hand, the total number of configurations can be represented by $\log_2 N_c = N(N-1)/2$ bits. Therefore, the efficiency of the system, η, is

$$\eta \equiv \log(N_s)/\log(N_c) \leq 1. \tag{3}$$

This quantity represents the ratio between the number of embedded bits in the space of the patterns and the number of embedded bits in the space of the weights. It is clear that η is bounded from above by one, since each configuration can contribute only one new symbol. Nevertheless, in the following we will explain the fact that even when $\eta \leqslant 1$, the capacity per bit can be greater than one.

No analytical method is known for the calculation of the efficiency η, and therefore in the following η is examined numerically. However, even numerically, the estimation of η is a hard combinatorial problem, since the number of symbols scales with 2^{N^2} and their size is bounded from above by order 2^N.

The efficiency of finite size classical systems of size N and with binary weights was examined numerically using the following heavy algorithm.

(a) Run over all the possible configurations, N_c.

(b) Find the symbol for the actual configuration, namely, a set of stable patterns. The symbol is defined to be of size l if it stands for l stable patterns.

(c) If the symbol is either equal or is a subset of an already existing symbol, then continue with the next configuration.

(d) If it is a new symbol of size l, for instance, add this symbol to the list of symbols and delete all symbols in this list of size $<l$ which are included in this symbol. Continue with the next configuration. Note that the maximal size of the symbol is taken to be 2^{N-1} rather than 2^N, since equation (1) obeys a global inversion symmetry.

At the end of this procedure the list of the symbols is the minimal list which characterizes the computational capability of a binary system of size N. The results for an even $N \leqslant 12$ are presented in table 1. Since the number of configurations grows exponentially with N^2, an exhaustive search over all the configurations is possible only for $N \leqslant 6$. For $N > 6$ only a subspace of the configurations was examined, either by fixing the value of part of the weights or by choosing random fraction of the configurations (fluctuations among different samples are found to be negligible). The number of examined configurations is denoted by N_c, where in an exhaustive search $N_c = 2^{(N-1)N/2}$, as was previously defined. It was found in the simulations that η is a non-decreasing function of N_c, the number of examined configurations, and hence the simulations indicate a lower bound for η. Furthermore, numerically it is found that when N_c increases from 2^{22} to 2^{24} for $N = 8$, for instance, the number of independent symbols, N_s, increases, but the change in the value of η (see equation 3) is negligible. These results indicate that the lower bound should be very close to the exact value of η which is obtained by an exhaustive search (2^{28} configurations in the discussed example). Details of the simulations and the behaviour of η as a function of N_c will be given elsewhere [6].

There are two remarkable results which one can conclude from the simulations for classical systems. The first conclusion is that η is an increasing function of N, where

Table 1. Number of symbols (symbols) and the efficiency (η) for classical systems of size N.

N	Symbols	η
2	2	1.000
4	8	0.300
6	896	0.654
8		~0.755
10		~0.820
12		~0.998

$N=2$ is an exceptional value. The second conclusion is that the histogram of the minimal list of symbols which characterizes the computational capability of the system has a wide distribution. For $N=2$ there are 2 symbols of size 1. For $N=4$ there are 8 symbols of size 6. For $N=6$ there are 480, 384, 32 symbols of the sizes 5, 6, 10 respectively, where the full list of the 2^{15} symbols (for each one of the configurations) is given by 17 952, 576, 7680, 480, 480, 384, 32 for the sizes 1, 2, 3, 4, 5, 6, 10, respectively. For $N=8$ and $N_c=2^{24}$, the histogram of the minimal list of symbols is given in table 2, where for $N=10$ and 12 the histogram will be presented elsewhere. Note, however, that the maximal symbol for $N=10$ and 12 is found to be 126 and 462, respectively. In general the maximal symbol is at least of size $N!/2(N/2)!^2$ and the number of such different symbols is 2^{N-1}. This result one can derive from the antiferromagnetic configuration and the 2^N configurations which are obtained from it by gauge transformations. It is also remarkable to note that all these histograms are characterized by the absence of symbols up to some minimal size and by a maximum of the distribution at a greater size. Furthermore, the result that the minimal list of symbols has a long tail plays an important role when the notions capacity and information are compared below.

Table 2. Number of symbols and their sizes for $N=8$ and $N_c=2^{24}$.

Size	Symbols	Size	Symbols
4	1 000	10	10 080
5	85 032	13	14 248
6	113 736	15	32
7	17 056	16	5 376
8	40 320	35	8

The efficiency of binary quantum spin systems to convert information from the presentation of the weights to the representation of the spins was examined in a similar way to the classical one with the following necessary modifications.

(*a*) The stable states are replaced by the eigenfunctions of the quantum spin Hamiltonian, equation (2).

(*b*) The size of the symbols is fixed and equal to 2^N, which is the available number of dimensions for N quantum $\frac{1}{2}$ spins.

(*c*) A system of odd number of spins is meaningful, since the stable states are well defined.

(*d*) The eigenfunctions of a configuration $\{J\}$ are the same as for a configuration $\{-J\}$. Therefore, η is bounded from above by $1-1/\log_2(N_c)$.

(*e*) Each symbol is constructed from degenerate subspaces $\{d_i\}$ such that $\Sigma\, d_i=2^N$. Each vector which is spanned by the subspace d_l, for instance, has the same eigenvalue E_l.

(*f*) A configuration contributes new information if, and only if, its symbol contributes new information which is not included in any of the other symbols. More precisely, a symbol A ($\Sigma\, A_i=2^N$) is included in a symbol B ($\Sigma\, B_j=2^N$) if, and only if, for any i, $A_i \subseteq$ in one of the subspaces B_j.

One might consider the following different definition for a new symbol. A symbol A does not contribute new information if, and only if, any subspace A_i is included in a subspace B_j of *any* other symbol, $A\neq B$. It is clear that this definition gives a shorter list of symbols. However, this definition is wrong, since the correlations among the embedded subspaces in each one of the symbols are lost. As an illustrative example, assume that a symbol A is constructed from the subspaces x and y, a symbol B is constructed from the subspaces x_1 and y_1 and a symbol C is constructed from the subspaces x_2 and y_2, where $x \subseteq x_1$ and $y \subseteq y_2$. According to this definition of a new symbol, a symbol A is included in symbols B and C, but the information that the subspaces x and y can be embedded *simultaneously* in one configuration is lost. Although this definition of a new symbol is wrong for estimating the embedded information in quantum systems, it is relevant for the estimation of the capacity of quantum systems, as is discussed below.

The results for the efficiency of finite quantum spin systems of the sizes $2 \leqslant N \leqslant 8$ are summarized in table 3. An exhaustive search over all configurations is possible only for $N \leqslant 6$. For $N = 7$ and 8 only a subspace of the configurations was examined, either by fixing the value of part of the weights or by choosing random fraction of the configurations. As for the classical systems, it is found that η is roughly independent of the fraction of examined configurations. Hence the estimated η should be very close to the exact value of η obtained in an exhaustive search. It is clear that the efficiency, η, is an *increasing* function of N, where $N = 4$ is an exceptional value which is due to strong fluctuations in a small system.

Table 3. Number of symbols (symbols) and the efficiency (η) for quantum systems of size N.

N	Symbols	η
2	1	0.000
3	1	0.000
4	32	0.833
5	282	0.814
6	12 772	0.909
7		~0.93
8		~0.96

The number of *different* subspaces which form the list of the minimal number of symbols was also calculated and shows that for $N = 2$ there are 2 subspaces of the sizes 1, 3 for $N = 3$ there are 2 subspaces of size 4, for $N = 4$ there are 12, 1, 1, 3 subspaces of the sizes 4, 7, 9, 11 and for $N = 5$ there are 1, 35, 1 symbols of the sizes 10, 12, 16, respectively. The list for $N = 6$ is given in table 4. In these calculations, the subspace A_k is counted only once even if it appears in many symbols and it is deleted if it is included in any subspace B_l. Basically, this is the minimal list of subspaces from which one can construct the minimal list of symbols which characterizes the computational ability of the system. As for the classical case, the distribution of the histogram of the subspaces is wide, and the average subspace increases as a function of N and is equal to 2, 4, 5.7, 12.05 and 14.5 for $N = 2$, 3, 4, 5 and 6, respectively. The detailed results for the cases $N = 7$ and 8 will be given elsewhere [7].

The minimal list of subspaces which describes the computational ability of a quantum system of size N has a long tail, as for classical systems. It is clear that the total number of the dimensions of this list of symbols can be even greater than the number of configurations, $2^{N(N-1)/2}$, and the average subspace can be greater than $N/2$. This is indeed the case for $N = 4$, 5 and 6. Hence, it is not surprising that one

Table 4. The minimal list of subspaces for $N=6$ quantum system.

Size	Symbols	Size	Symbols
4	720	18	180
5	120	19	41
7	7	20	90
8	360	21	600
10	780	22	270
11	360	23	225
12	270	24	120
13	420	25	202
14	255	27	46
15	600	29	195
16	330	30	15
17	180	31	10

can define similar quantities to the capacity per bit whose measurements give results which are greater than 1, as discussed below for classical systems.

In the following, the surprising result that the capacity per bit can be greater than 1 is explained, using the consistent information theory examined above.

In the case that the histogram of the minimal list of symbols is constructed only from one size and its logarithm does not scale with N, the capacity per bit should be less than or equal to 1, as is discussed below. Since the number of symbols is bounded by the number of configurations, than in the discussed case the maximal number of symbols of size P should be less than the total number of configurations

$$\binom{2^N}{P} \leqslant 2^{N(N-1)/2}. \tag{4}$$

Therefore, in the leading order $P \leqslant N/2$ and the capacity per bit is obviously less than or equal to 1. However, in the case that the distribution of the minimal list of symbols has a long tail, a symbol of size P_1 contributes $P_1!/(P_1-P)!P!$ subsymbols of size P. Neglecting the partial overlap among the symbols, an upper bound for the capacity is given by the maximal P which obeys the following inequality

$$\sum_{l \geqslant P} n_l \binom{l}{P} \leqslant \binom{2^N}{P} \simeq 2^{NP} \tag{5}$$

where n_l is the number of symbols of size l and $\Sigma\, n_l \leqslant 2^{N_c}$ for the classical case. In this case the capacity per bit can be greater or less than 1 depending on the details of the distribution of the sizes of the symbols. One can say that, in general, the measure of the capacity is not the 'natural language' of the system. Indeed, one can verify from the abovementioned results that equation (5) gives a capacity per bit greater than one for $N \leqslant 8$, which is possible due to the wide distribution of the size of the symbols. This fact was also confirmed by direct simulations to fix the capacity [7], and by an exact calculation of the left side of equation (5) where the correlations among the symbols were taken into account.

In the case of feedforward networks, the size of the symbols (boolean functions) is fixed to be 2^{N_i}, where N_i is the size of the input. Each symbol stands for the type and the number of embedded relations between the input and the output for a particular configuration. Hence, the capacity per bit is less than 1. For the binary perceptron [8], for instance, the capacity is fixed by the maximal P which obeys the inequality

$$2^N \binom{2^N}{P} \geqslant \binom{2^{N+1}}{P}. \qquad (6)$$

The left side of equation (6) stands for the maximal number of symbols of size P which can be constructed from 2^N symbols where the size of each one of them is also 2^N. The right side is the total number of symbols of size P. From the solution of equation (6) one obtains that $P \leqslant N$.

Finally, note that the behaviour of the system with a deeper synaptic depth is still in question. Nevertheless, it is expected that in such a case η is higher in the quantum case than in the classical one, since the space of the particles in the quantum case is continuous.

One of us (IK) thanks Professor J A Wheeler for his encouragement and for bringing [6] to his attention.

References

[1] For review, see Amit D J 1989 *Modeling Brain Function* (New York: Cambridge Univ. Press)
[2] See, for intance, the memorial volume of Elizabeth Gardner 1989 *J. Phys. A: Math. Gen.* **22**
[3] Barkai E, Hansel D and Kanter I 1990 *Phys. Rev. Lett.* **65** 2312
[4] Barkai E and Kanter I 1991 *Europhys. Lett.* **14** 107
[5] Kanter I 1992 *Phys. Rev.* A **15** 6051
[6] Pierce J R 1962 *Symbols, Signals and Noise* (London: Hutchinson Science Library)
[7] Kanter I and Eisenstein E unpublished
[8] Minsky M L and Paper S 1969 *Perceptron* (Cambridge, MA: MIT Press)