On the center of mass of Ising vectors

EUROPHYSICS LETTERS 15 July 1999

Europhys. Lett., **47** (2), pp. 139-144 (1999)

On the center of mass of Ising vectors

M. Copelli, M. Bouten, C. Van den Broeck and B. Van Rompaey

Limburgs Universitair Centrum - B-3590 Diepenbeek, Belgium

(received 18 January 1999; accepted in final form 19 May 1999)

PACS. 02.50−r – Probability theory, stochastic processes, and statistics.
PACS. 87.10+e – Biological and medical physics: General theory and mathematical aspects.
PACS. 64.60Cn – Order-disorder transformations; statistical mechanics of model systems.

Abstract. – We show that the center of mass of Ising vectors that obey some simple constraints, is again an Ising vector.

Many problems in statistical mechanics are formulated in terms of an N-dimensional vector $\mathbf{J}$, with components $J_i, i = 1, ..., N$ that take only binary values $J_i = \pm 1$. We will call such a vector an Ising vector. Its components represent for example a spin state (Ising model [1]), the occupancy of a site, etc. In the thermodynamic limit $N \to \infty$, only a subset of all the possible configurations $\{\mathbf{J}\}$ are typically realized. In many cases, they are characterized by simple constraints of the form

$$\lim_{N\to\infty} \frac{\mathbf{J}\cdot\mathbf{B}}{N} = R\,, \qquad \lim_{N\to\infty} \frac{\mathbf{J}\cdot\mathbf{J}'}{N} = q, \tag{1}$$

where $\mathbf{J}$ and $\mathbf{J}'$ are typical members of the subset, $\mathbf{B}$ is a symmetry-breaking direction (imposed from the outside or arising through a phase transition), while q and R are physical properties describing the resulting macroscopic state (for example the magnetization or the density). In this letter, we focus on the center of mass of the vectors $\mathbf{J}$, that satisfy the above constraints. We report the surprising finding that it is an Ising vector whenever $\mathbf{B}$ is Ising.

To construct the center of mass we follow a Monte Carlo approach by choosing at random n vectors $\mathbf{J}^a, a = 1, ..., n$, that satisfy the constraints, and considering their center of mass:

$$\mathbf{Y} = C^{-1} \sum_a \mathbf{J}^a, \tag{2}$$

with the proportionality factor $C = \sqrt{n + n(n-1)q}$, so that the normalization condition $\mathbf{Y}^2 = N$ is obeyed. In general the vector $\mathbf{Y}$ is not Ising, but our contention is that it becomes Ising in the limit $n \to \infty$, provided $\mathbf{B}$ is Ising. In view of the permutation symmetry between the coordinate axes, it will be sufficient to prove that $B_1 Y_1 = C^{-1} \sum_a B_1\ J_1^a$ only takes the values $+1$ and -1 in this limit. To show that this is the case, we focus our attention on the probability density $P(y)$ of the variable $y = n^{-1}\sum_a B_1\ J_1^a$, which differs by a factor $n^{-1}\sqrt{n+n(n-1)q} \overset{n\to\infty}{\to} \sqrt{q}$ from $B_1 Y_1$. It is given by

$$P(y) \sim \int \left[\prod_a^n \mathrm{d}\mathbf{J}^a P_b(\mathbf{J}^a)\ \delta(\mathbf{J}^a\cdot\mathbf{B} - NR)\right] \times \\ \times \left[\prod_{a<b} \delta(\mathbf{J}^a\cdot\mathbf{J}^b - Nq)\right] \delta\left(y - n^{-1}\sum_a B_1\ J_1^a\right), \tag{3}$$

where P_b is the measure restricting to vectors with binary components,

$$P_b(\mathbf{J}) = \prod_{j=1}^N \left[\frac{1}{2}\delta(J_j - 1) + \frac{1}{2}\delta(J_j + 1)\right], \tag{4}$$

and the proportionality constant has to be determined from the normalization condition $\int_{-\infty}^{\infty} \mathrm{d}y\ P(y) = 1$. The r.h.s. of (3) resembles an ordinary replica calculation [2], but with as limit of interest the number of replicas n tending to infinity.

Rather than following the standard but lengthy calculations that are usual in this case, we present a more elegant, direct and expedient procedure. Since $y = n^{-1}\sum_a x_a$, with $x_a = J_1^a B_1$, we evaluate the joint probability density $P(\mathbf{x})$ of the n-dimensional vector with binary components x_a, $a = 1, \dots, n$. Since all choices of the vectors $\mathbf{J}^a$ that satisfy the constraints can be realized, the Shannon entropy is maximized under the constraints (1). Hence $P(\mathbf{x})$ is found by maximizing [3] its Shannon entropy $-\sum_{\mathbf{x}} P(\mathbf{x}) \ln P(\mathbf{x})$, subject to the constraints

$$\langle x_a \rangle = \frac{\mathbf{J}^a\cdot\mathbf{B}}{N} = R, \\ \langle x_a x_b \rangle = \frac{\mathbf{J}^a\cdot\mathbf{J}^b}{N} = q, \qquad (a<b). \tag{5}$$

One finds

$$P(\mathbf{x}) = Z^{-1} \exp\left[\sum_a \hat{R}_a x_a + \sum_{a<b} \hat{q}_{ab} x_a x_b\right], \tag{6}$$

where Z follows from the normalization of $P(\mathbf{x})$. The values of the Lagrange multipliers $\{\hat{R}_a\}$ and $\{\hat{q}_{ab}\}$ have to be determined from the constraints (5). In view of the permutation symmetry in the replica indices, $\hat{R}_a$ and $\hat{q}_{ab}$ must be independent of a and b, $\hat{R}_a = \hat{R}$, $\hat{q}_{ab} = \hat{q}$, rendering the evaluation of Z very simple:

$$Z(\hat{R}, \hat{q}) = e^{-n\hat{q}/2} \int Dz \left[\cosh\left(\hat{R} + z\sqrt{\hat{q}}\right)\right]^n, \tag{7}$$

while (5), determining $\hat{R}$ and $\hat{q}$, reduce to

$$R = \frac{1}{n}\frac{\partial}{\partial \hat{R}} \ln Z = \frac{\int \mathrm{d}u\ \exp\left[-(u-\hat{R})^2/2\hat{q}\right] (\cosh u)^n\ \tanh u}{\int \mathrm{d}u\ \exp\left[-(u-\hat{R})^2/2\hat{q}\right] (\cosh u)^n}, \\ q = \frac{2}{n(n-1)}\frac{\partial}{\partial \hat{q}} \ln Z = \frac{\int \mathrm{d}u\ \exp\left[-(u-\hat{R})^2/2\hat{q}\right] (\cosh u)^n\ \tanh^2 u}{\int \mathrm{d}u\ \exp\left[-(u-\hat{R})^2/2\hat{q}\right] (\cosh u)^n}. \tag{8}$$

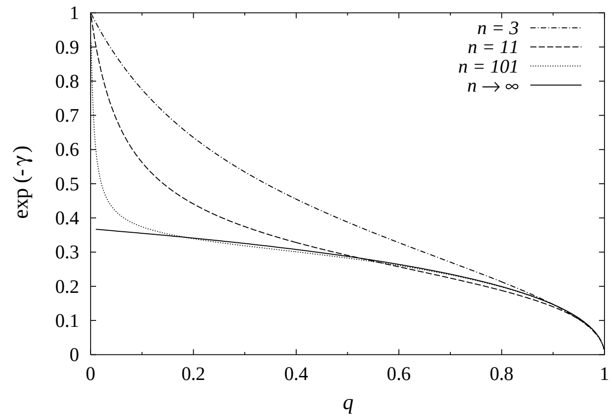

Fig. 1. – In order to account for a logarithmic divergence in the limit $q \to 1$, $\exp[-\gamma], \gamma = n\hat{q}$, is plotted as a function of q for several values of n.

As a result of the "replica symmetry", we conclude from (6) that $P(\mathbf{x})$ is in fact a function of $\sum_a x_a = ny$. Hence $P(y)$ is obtained from $P(\mathbf{x})$ by multiplication with a combinatorial factor, expressing the freedom to choose which $n(1+y)/2$ components of $\mathbf{x}$ are $+1$ and which remaining ones are -1, for a given total value ny of their sum:

$$P(y) \sim \binom{n}{\frac{n(1+y)}{2}} \exp\left[n\hat{R}y + \frac{n^2\hat{q}y^2}{2}\right] . \tag{9}$$

y can take the values $-1, -1+2/n, \ldots, 1-2/n, 1$, and the proportionality constant is again fixed by normalization. This result is in agreement with a direct evaluation of (3) but very different from the result for continuous components discussed in [4]. Unfortunately, the above expression is quite complicated, especially in view of the fact that we did not succeed in solving explicitly eqs. (8) determining the Lagrange multipliers. Concordantly, the components of $\mathbf{Y}$ are not binary for any finite n. As an illustration, we have included in fig. 1 the results obtained by a numerical solution of (8) for the special case $q = R$ and several values of n. The corresponding results for the probability density for y (or equivalently, B_1Y_1), are plotted in fig. 2.

In order to extract the asymptotic behavior for the $n \to \infty$ limit, one needs to guess the asymptotic dependence on n of the Lagrange parameters. The correct scaling appears quite naturally in the calculations for the simpler case of vectors $\mathbf{J}$ with continuous components. Here, we just note by inspection that eqs. (8) for the properly scaled Lagrange parameters $\rho \equiv n\,\hat{R}$ and $\gamma \equiv n\,\hat{q}$ read

$$\begin{aligned} R &= \frac{\int \mathrm{d}u\, e^{-n\phi(u)} \sinh(u\rho/\gamma) \tanh u}{\int \mathrm{d}u\, e^{-n\phi(u)} \cosh(u\rho/\gamma)} , \\ q &= \frac{\int \mathrm{d}u\, e^{-n\phi(u)} \cosh(u\rho/\gamma) \tanh^2 u}{\int \mathrm{d}u\, e^{-n\phi(u)} \cosh(u\rho/\gamma)} , \end{aligned} \tag{10}$$

where

$$\phi(u) \equiv \frac{u^2}{2\gamma} - \ln\cosh u . \tag{11}$$

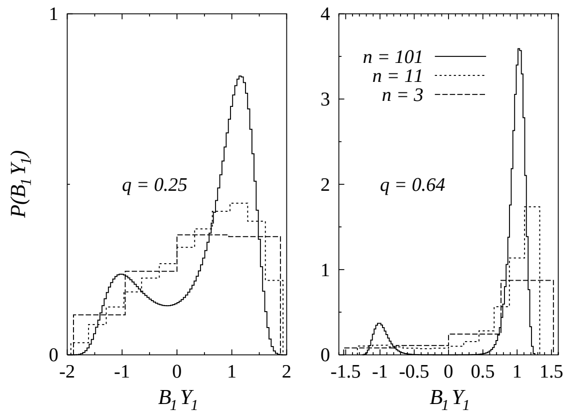

Fig. 2. – Probability density for B_1Y_1 according to eq. (9) for $q = R$ and several values of n. The legend is the same for both plots.

The appearance of the hyperbolic functions of $u\rho/\gamma$ in eqs. (10) is due to the fact that ϕ is even. The saddle point approximation can now, for $n \to \infty$, be applied in a straightforward manner on the u-integrations, leading to the following simple and explicit solutions for the scaled Lagrange variables:

$$\gamma = \frac{\mathrm{arctanh}\sqrt{q}}{\sqrt{q}} \, ,$$
$$\rho = \frac{\mathrm{arctanh}(R/\sqrt{q})}{\sqrt{q}} \, . \tag{12}$$

Inserting this result together with the asymptotic expression for the combinatorial factor in (9), one finally obtains the following asymptotic result for $P(y)$:

$$\begin{aligned} P(y) \quad &\sim \quad \exp[\rho\, y] \exp n \left[\gamma\, y^2/2 - \ln\sqrt{1-y^2} - y \,\mathrm{arctanh}\, y\right] \\ &\overset{n\to\infty}{\to} \frac{1}{2}\left(1+\frac{R}{\sqrt{q}}\right)\delta(y-\sqrt{q}) + \frac{1}{2}\left(1-\frac{R}{\sqrt{q}}\right)\delta(y+\sqrt{q}) \, . \end{aligned} \tag{13}$$

In view of the aforementioned relation between y and the (first) component of $\mathbf{Y}$, the convergence of the latter to an Ising vector follows immediately.

As a first application of the above result, we turn to the case of an Ising spin system in the ferromagnetic phase. Choosing the vector $\mathbf{B}$ with all its components equal to 1, we note that R plays the role of the magnetization. All the spin states $\mathbf{J}$ are otherwise allowed, and lie on the rim of the N-dimensional sphere with radius $\sqrt{N}$ at fixed angle $\arccos R$ with $\mathbf{B}$. It is thus clear that the center of mass is $\mathbf{B}$ itself. This trivial result is recovered from (13) by noting that $q = R^2$ in this case. Note that the constraint $\mathbf{J}^a \cdot \mathbf{J}^b = Nq$ is therefore redundant, which implies $\hat{q} = 0$ and $\hat{R} = \mathrm{arctanh} R$ (a result valid for any n).

A case of special symmetry is $q = R$. In this scenario, no macroscopic measure allows to distinguish between the symmetry-breaking direction $\mathbf{B}$ and each of the vectors $\mathbf{J}^a$. The Lagrange parameters also present the symmetry $\hat{R} = \hat{q}$, which can be seen from eqs. (8).

A third case of interest is the limit $R \to 0$, while q remains finite. In this case, the $\mathbf{J}$-vectors lie in the subspace orthogonal to $\mathbf{B}$ and satisfy as single constraint the prescribed mutual overlap q. From eqs. (8) it is clear that $\hat{R} = 0$ is automatically satisfied, and one concludes

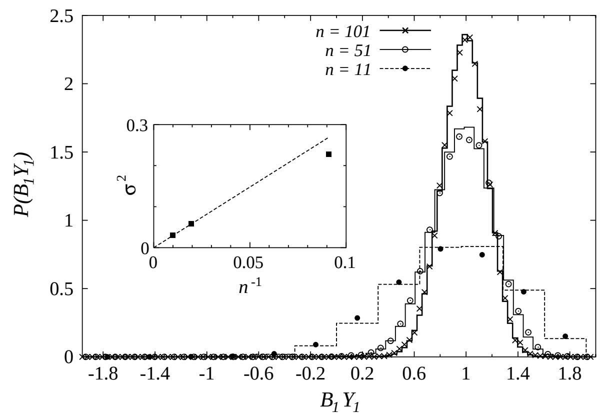

Fig. 3. – Probability density for B_1Y_1 in the $q = R^2$ scenario and several values of n. The lines represent the theoretical curve (9), while points represent results from a simulation of the mean-field ferromagnetic Ising model with $N = 100$, dimensionless temperature 1.09 and dimensionless magnetic field 0.1 (amounting to a magnetization $R \simeq 0.5$—see text for details). Inset: second cumulant (σ^2) of the distribution as a function of $1/n$. The squares represent the simulations, while the dashed straight line corresponds to the theory (in the asymptotic limit $n \to \infty$, eq. (13) implies $\sigma^2 \sim (1-R^2)/(nR^2)$).

from (13) that the center of mass of the **J**-vectors is again an Ising vector (but its components are equally likely to be $+1$ or -1).

It is very tempting to apply the results above to neural network learning problems [5] where a student perceptron **J** learns from examples generated by a teacher perceptron **B** [6-8]. Indeed, the so-called Gibbs learning [9] presents the symmetry $q = R$, and the interest of the center of mass is that it is, according to a simple general argument [10], see also [11], the "best" student having the largest overlap with the teacher (namely $\sqrt{R}$). Accordingly, $R = 0$ and $q \neq 0$ are constraints satisfied by Ising vectors which solve the capacity problem [12]. However, these are disordered systems, and the conditions on R and q alone do not convey all the information which is necessary to describe the constraints in **J** space. Therefore, even though the constraints (1) are satisfied in neural network problems, result (13) does not apply to them.

We conclude with a verification of the theoretical prediction (9) by running simulations for the mean-field ferromagnetic Ising spin model, fig. 3. The Metropolis algorithm was allowed to run for a number of Monte Carlo steps per site (MCS) until thermalization was considered to be achieved. Then vectors were sampled every 5 MCS (to allow sufficient decorrelation between consecutive samplers) and summed to construct the center of mass. The small discrepancy for $N = 100$ with the theoretical prediction is due to finite-size effects. For $N = 1000$, the results are nearly indistinguishable on the scale of the figure from the theoretical values. It is interesting to note that the hard constraints of eq. (1) are satisfied *only* in the thermodynamic limit. In the simulations R (and q) are distributed with peaks whose width scales with $N^{-1/2}$. Nonetheless the effect of these fluctuations on the resulting $P(B_1Y_1)$ is negligible.

***

The authors would like to thank the referees for useful suggestions. We also acknowledge support from the FWO Vlaanderen and the Belgian IUAP program (Prime Minister's Office).

REFERENCES

[1] Ising E., *Z. Phys.*, **31** (1925) 253.
[2] Mézard M., Parisi G. and Virasoro M. A., *Spin Glass Theory and Beyond* (World Scientific, Singapore) 1987.
[3] Jaynes E. T., in *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, edited by R. D. Rosenkrantz (D. Reidel Publishing Company) 1983, pp. 39-76.
[4] Schietse J., Bouten M. and Van den Broeck C., *Europhys. Lett.*, **32** (1995) 279.
[5] Hertz A. J., Krogh A. and Palmer R. G., *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA) 1991.
[6] Gardner E. and Derrida B., *J. Phys. A*, **22** (1989) 1983.
[7] Györgyi G. and Tishby N., in *Neural Networks and Spin Glasses*, edited by W. K. Theumann and R. Köberle (World Scientific, Singapore) 1990, pp. 3-36.
[8] Seung H. S., Sompolinsky H. and Tishby N., *Phys. Rev. A*, **45** (1992) 6056.
[9] Györgyi G., *Phys. Rev. A*, **41** (1990) 7097.
[10] Watkin T. L. H., *Europhys. Lett.*, **21** (1993) 871.
[11] Opper M. and Haussler D., *Phys. Rev. Lett.*, **66** (1991) 2677.
[12] Gardner E. and Derrida B., *J. Phys A*, **21** (1988) 271.