

# Inverse spin glass and related maximum entropy problems

Michele Castellana<sup>1</sup> and William Bialek<sup>1,2</sup>

<sup>1</sup>*Joseph Henry Laboratories of Physics and Lewis–Sigler Institute for Integrative Genomics,  
Princeton University, Princeton, New Jersey 08544*

<sup>2</sup>*Initiative for the Theoretical Sciences, The Graduate Center,  
City University of New York, 365 Fifth Ave., New York, New York 10016*

(Dated: September 10, 2014)

If we have a system of binary variables and we measure the pairwise correlations among these variables, then the least structured or maximum entropy model for their joint distribution is an Ising model with pairwise interactions among the spins. Here we consider inhomogeneous systems in which we constrain (for example) not the full matrix of correlations, but only the distribution from which these correlations are drawn. In this sense, what we have constructed is an inverse spin glass: rather than choosing coupling constants at random from a distribution and calculating correlations, we choose the correlations from a distribution and infer the coupling constants. We argue that such models generate a block structure in the space of couplings, which provides an explicit solution of the inverse problem. This allows us to generate a phase diagram in the space of (measurable) moments of the distribution of correlations. We expect that these ideas will be most useful in building models for systems that are nonequilibrium statistical mechanics problems, such as networks of real neurons.

PACS numbers: 05.20.-y, 02.50.Tt, 87.10.-e

Systems at thermal equilibrium are in a state of maximum entropy. But maximizing entropy also provides a method for building models of systems, whether in equilibrium or not, that are consistent with some set of measurements but otherwise have as little structure as possible [1]. Concretely, we consider a system described by variables  $\sigma \equiv \{\sigma_1, \sigma_2, \dots, \sigma_N\}$ , and we would like to construct the probability distribution  $P(\sigma)$  over these states. We can take from experiment measurements on the expectation values of various operators  $O_1(\sigma), O_2(\sigma), \dots, O_K(\sigma)$ , and so we insist that

$$\sum_{\sigma} P(\sigma) O_{\mu}(\sigma) = \langle O_{\mu}(\sigma) \rangle_{\text{expt}}. \quad (1)$$

Searching all probability distributions that obey these constraints, we can find the one which has the maximum entropy, and the result is a Boltzmann-like distribution  $P(\sigma) = e^{-E(\sigma)} / Z(\{g_{\mu}\})$ , with an effective energy  $E(\sigma) = \sum_{\mu=1}^K g_{\mu} O_{\mu}(\sigma)$ , where  $Z(\{g_{\mu}\})$  is the partition function enforcing the normalization of  $P(\sigma)$ . To complete the construction we must find the values of the coupling constants  $g_{\mu}$  that satisfy the constraints in Eq. (1). This is the inverse of the usual problem in statistical mechanics: rather than knowing the coupling constants and trying to predict expectation values, we are given the expectation values and must determine the coupling constants. In general this inverse problem is hard, and application of the maximum entropy method to real systems usually depends on detailed numerics.

Recent applications of the maximum entropy approach to a wide variety of biological systems—patterns of activity in networks of neurons [2–11], the structure and dynamics of biochemical and genetic networks [12, 13], the ensemble of amino acid sequences in families of proteins [14–20], and ordering in flocks of birds [21, 22]—have generated renewed interest in the inverse problem. A variety of approximations and algorithmic solutions have been suggested, based on methods borrowed from statistical mechanics [23, 24], statistical inference [25, 26] and ma-

chine learning [27]. The essential difficulty is that these systems are strongly inhomogeneous. As an example, if the  $\langle \hat{O}_{\mu}(\sigma) \rangle_{\text{expt}}$  are the correlations between the spikes generated by pairs of neurons in a network, in principle we have a correlation matrix with an arbitrary structure and hence  $N(N-1)/2$  coupling constants  $\{g_{\mu}\}$  that need not have any simple relation to one another. In this setting even the forward problem ( $\{g_{\mu}\} \rightarrow \{\langle O_{\mu}(\sigma) \rangle\}$ ) is difficult.

One of the lessons from the statistical mechanics of disordered systems is that we can make statements about an ensemble of systems with randomly chosen parameters even if it is difficult to solve the problem of a single system with inhomogeneous parameters [28]. Here we apply this lesson to the inverse problem. Suppose that  $\mu$  is a local index referring (for example) to single sites or links in a network: rather than asking for the expectation value of each local operator, we will ask about the distribution of expectation values across the network. This idea is guided by previous works on maximum entropy models for neural activity, where one considers ensembles of networks constructed by drawing mean spike probabilities and pairwise correlations from the observed distribution of these quantities across a real network, and then solves the full inverse problem for many members of this ensemble [4, 7]. Interestingly, these “typical” networks have many properties in common with the real network. The advance here is that the system is insensitive to the precise topology of the network, and the physical information is encoded into the distribution of expectation values of local operators across the network rather than in the expectation values of all local operators: this will be the working hypothesis of the maximum entropy approach presented in this Letter. More concretely, given the moments  $M_n = \frac{1}{K} \sum_{\mu=1}^K (\langle O_{\mu}(\sigma) \rangle_{\text{expt}})^n$  for  $n = 1, 2, \dots, R$ , we will show an analytic approach to construct the probability distribution over states  $P(\sigma)$  that is consistent with these moments, but otherwise as random as possible.

To maximize the entropy of  $P(\sigma)$  subject to constraints

on the moments  $\{M_n\}$ , we proceed as usual by introducing Lagrange multipliers, so that we should maximize

$$\mathcal{L} = - \sum_{\sigma} P(\sigma) \ln P(\sigma) + N\lambda_0 \left[ \sum_{\sigma} P(\sigma) - 1 \right] + \quad (2)$$

$$+ N \sum_{n=1}^R \lambda_n \left[ \frac{1}{K} \sum_{\mu=1}^K \left( \sum_{\sigma} P(\sigma) O_{\mu}(\sigma) \right)^n - M_n \right],$$

where  $\lambda_0$  enforces normalization, and we keep the first  $R$  moments. Notice that the first term in Eq. (2) is extensive; explicit factors of  $N$  insure that the other terms also are extensive. Solving  $\partial\mathcal{L}/\partial P(\sigma) = 0$ , we find that the maximum entropy distribution is again a Boltzmann distribution, but with the coupling constants related, self-consistently, to the expectation values:

$$P(\sigma) = \frac{1}{Z(\{g_{\mu}\})} \exp \left[ \sum_{\mu=1}^K g_{\mu} O_{\mu}(\sigma) \right], \quad (3)$$

where  $g_{\mu} \equiv \frac{N}{K} \sum_{n=1}^R n \lambda_n \phi_{\mu}^{n-1}$ , and  $\phi_{\mu} = \sum_{\sigma} P(\sigma) O_{\mu}(\sigma)$  is the expectation value of  $O_{\mu}(\sigma)$  in the distribution  $P(\sigma)$ ; we still must adjust the  $\{\lambda_n\}$  to match the observed  $\{M_n\}$ .

In the simplest version of these ideas, the variables  $\sigma_i$  are Ising spins, and the operators  $O_{\mu}(\{\sigma\}) = \sigma_{\mu}$  are the individual spins themselves (hence  $K = N$ ). The maximum entropy model consistent with knowing the expectation values of every individual spin corresponds to a collection of independent spins in local magnetic fields

$$P(\sigma) = \frac{1}{Z(\{h_i\})} \exp \left( \sum_{i=1}^N h_i \sigma_i \right), \quad (4)$$

with  $\langle \sigma_i \rangle = \tanh h_i$ , as usual. What happens if we know only a limited set of moments of the distribution of  $\langle \sigma_i \rangle$  across the system? For example, if we know only the first two moments  $m_1 \equiv \frac{1}{N} \sum_{i=1}^N \langle \sigma_i \rangle$ ,  $m_2 \equiv \frac{1}{N} \sum_{i=1}^N \langle \sigma_i \rangle^2$ , then the definition of  $g_{\mu}$  gives us

$$h_i = \lambda_1 + 2\lambda_2 \langle \sigma_i \rangle = \lambda_1 + 2\lambda_2 \tanh h_i. \quad (5)$$

For a given  $\lambda_1, \lambda_2$ , Eq. (5) has only a discrete set of solutions for  $h_i$ . Thus, the maximum entropy model consistent with the mean and variance of the magnetization across an ensemble of spins consists of independent spins in local magnetic fields which can take only discrete values. As we constrain more and more moments, the analog of Eq. (5) becomes a higher and higher order polynomial in  $\tanh h_i$ , and hence the number of discrete values of  $h_i$  increases, approaching a continuous distribution in the limit that we know all the moments.

An illustration of these ideas is shown in Fig. 1. We choose the expectation values  $\langle \sigma_i \rangle$  from the distribution shown in the left inset, and build maximum entropy models that are consistent either with knowledge of this full distribution or with just its first two moments. If the full distribution (all moments) are known, the model has a continuous distribution of fields, and we can compute the resulting maximal entropy, which we denote by  $s_{\text{all}}$  (see the Supplemental Material). Fixing just the first two

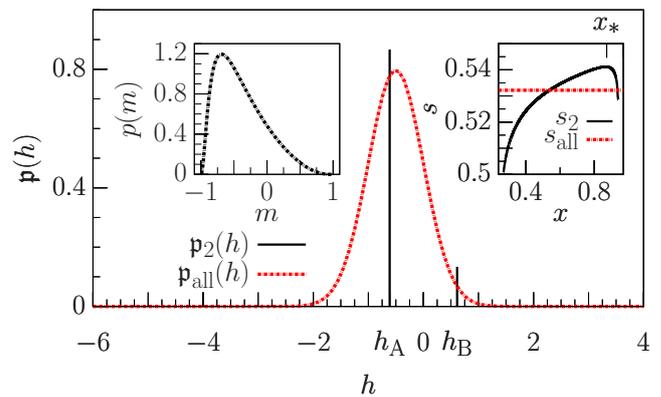


FIG. 1. Maximum entropy for independent spins. The distribution of local magnetization is shown in the left inset. In the main panel we show the field distributions  $p_2(h)$  (in black) and  $p_{\text{all}}(h)$  (in red) from the maximum entropy solution with two moments and all moments of the local magnetization, respectively. The distribution  $p_2(h)$  is given by two delta peaks at  $h = h_A$  and  $h = h_B$ . In the right inset we show the entropy per spin  $s_2$  (in black) in the two moment case vs. the fraction  $x$  of spins in group A, compared with the entropy in the all moment case  $s_{\text{all}}$  (in red). The optimal value  $x_*$  is also marked.

moments, we assume that there are two groups of spins A and B, with two discrete values of the field  $h_A$  and  $h_B$  acting on each group, and thus two values  $m_A, m_B$  of the local magnetizations  $\langle \sigma_i \rangle$ . Given a fraction  $x$  of spins in group A, we determine  $m_A, m_B$  by matching the first two moments

$$m_1 = x m_A + (1-x)m_B, \quad m_2 = x m_A^2 + (1-x)m_B^2, \quad (6)$$

we plug the solution into the equation  $\langle \sigma_i \rangle = \tanh h_i$ , and we solve for  $\lambda_1, \lambda_2$  by using Eq. (5): as a result, the entropy  $s_2$  depends only on the spin fraction  $x$ , and we fix  $x$  by maximizing  $s_2$ . It can be shown that this two block ansatz is exact. Indeed, we can fix  $\lambda_1, \lambda_2$  so that there are three distinct solutions, and the entropy depends on  $\lambda_1, \lambda_2$  only: we then maximize the entropy as a function of  $\lambda_1, \lambda_2$ , and at the maximum the fraction of spins  $\sigma_i$  with local field  $h_i$  equal to the third solution is equal to zero, and we are left with two values of the local fields (see the Supplemental Material). Importantly, Fig. 1 shows that a weakly constrained, random choice of magnetizations leads to a highly structured bimodal distribution of fields, even though we maximize the entropy and thus minimize structure in the distribution of spin configurations. Given that magnetizations and magnetic fields are related by the identity  $\langle \sigma_i \rangle = \tanh h_i$ , if the width of the magnetization distribution is increased—i.e. the standard deviation gets much larger than the mean—the distribution of fields tends to a bimodal distribution composed of two distant peaks rather than to a smooth distribution with a large width.

Maximum entropy models are much richer when the operators  $O_{\mu}$  live on the links between elements in a network rather than on nodes. Let us consider, then, the case where  $\mu = (i,j)$  denotes a pair of spins, and the

operators  $O_\mu(\{\sigma\}) \equiv \sigma_i \sigma_j$ . The number of operators  $K$  is the number of distinct pairs,  $N_p = N(N-1)/2$ , and we write  $C_{ij} = \langle \sigma_i \sigma_j \rangle$ . In what follows, we constrain the first two moments of the correlation distribution,  $C_1 = \frac{1}{N_p} \sum_{i>j} C_{ij}$ ,  $C_2 = \frac{1}{N_p} \sum_{i>j} C_{ij}^2$ . Equation (3) then becomes

$$P(\sigma) = \frac{1}{Z} \exp\left(\frac{N}{2N_p} \sum_{i,j} J_{ij} \sigma_i \sigma_j\right), \quad (7)$$

$$J_{ij} = \lambda_1 + 2\lambda_2 C_{ij}, \quad (8)$$

where in Eq. (7) we have incorporated a diagonal term with  $i = j$ , which is independent of  $\sigma$ , and the Lagrange multipliers are set by matching the moments of expectation values

$$C_1 = -\frac{\partial f}{\partial \lambda_1} - \frac{N}{2N_p}, \quad C_2 = -\frac{1}{2} \frac{\partial f}{\partial \lambda_2} - \frac{N}{2N_p} C_1, \quad (9)$$

where the free energy per spin is  $f = -(\ln Z)/N$ . Thus, the system is an Ising model in which the spin-spin couplings  $J_{ij}$  are related, bond by bond, to the spin-spin correlations  $C_{ij}$ . As with Eq. (5), it is difficult to imagine how the self-consistency condition in Eq. (8) can be satisfied by a broad distribution of couplings  $J_{ij}$ . In a system that is highly interconnected, the correlations between any pair of spins are dominated by the presence of multiple indirect paths, so that the  $C_{ij}$  need not even be correlated with the corresponding direct interactions  $J_{ij}$ . How then can Eq. (8) be solved? As with the case of independent spins, we suspect that the self-consistency condition in Eq. (8) can be satisfied only if the system breaks into blocks. If there are only a discrete set of possible  $J_{ij}$ , it seems possible that there will be only a discrete set of  $C_{ij}$ , and that we can arrange the pairs so that  $C_{ij}$  and  $J_{ij}$  are related linearly. With  $\lambda_1$  and  $\lambda_2$  fixed, we have done numerical experiments on systems of up to  $N = 16$  spins, solving Eq. (8) for the variables  $\{J_{ij}\}$ , and we have found that the couplings  $J_{ij}$  are driven to consist of two discrete values (see the Supplemental Material).

Guided by our numerical experiments, and by the case of independent spins above, we try a block ansatz: we divide the spins  $\{\sigma_i\}$  into two blocks,  $A \equiv \{\sigma_1, \dots, \sigma_{N_A}\}$ , and  $B \equiv \{\sigma_{N_A+1}, \dots, \sigma_N\}$ , and we assume that correlations between spins within a block take the value  $C_I$  while correlations between spins in different blocks take the value  $C_{II}$ ;  $x = N_A/N$  is the fraction of spins in block A. The parameters  $C_I, C_{II}$  are related to the moments by

$$\begin{aligned} C_1 &= [x^2 + (1-x)^2]C_I + 2x(1-x)C_{II}, \\ C_2 &= [x^2 + (1-x)^2]C_I^2 + 2x(1-x)C_{II}^2. \end{aligned} \quad (10)$$

The value of  $x$  will be set at the end of the calculation by maximizing the entropy, as above. It can be shown (see the Supplemental Material) that if the correlations satisfy the high temperature scaling  $C_1 = \mathcal{O}(1/N)$ ,  $C_2 = \mathcal{O}(1/N^2)$ , the entropy per spin is

$$s = \ln 2 - \frac{1}{N} \left[ \frac{1}{2} \ln \mathcal{D} + \lambda_1 (NC_1 + 1) + 2\lambda_2 (NC_2 + C_1) \right], \quad (11)$$

where  $\mathcal{D} = \det(1 - 2 \text{diag}(x, 1-x)^{-1} \cdot \mathcal{M})$ ,  $\text{diag}(x, 1-x)$  is a  $2 \times 2$  diagonal matrix with diagonal entries  $x, 1-x$ ,

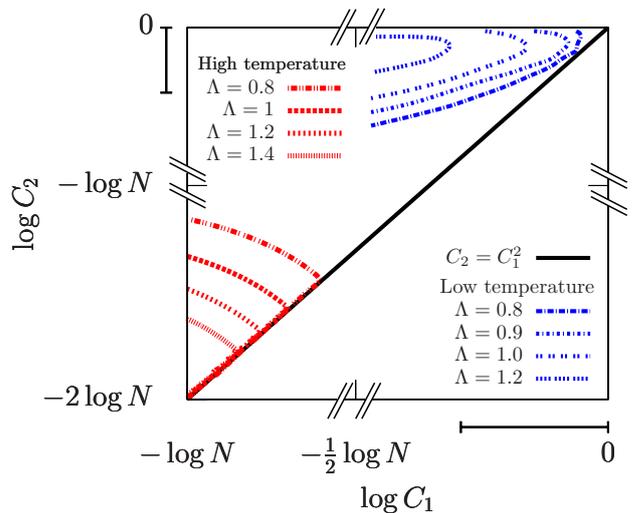


FIG. 2. Contour plot of the smallest eigenvalue  $\Lambda$  of the free energy Hessian as a function of  $\ln C_1, \ln C_2$  for the correlated spin case in the allowed region  $C_2 \geq C_1^2$ . There is a high temperature phase (in red) where  $C_1 = \mathcal{O}(1/N)$ ,  $C_2 = \mathcal{O}(1/N^2)$ , a low-temperature phase (in blue) where  $C_1 = \mathcal{O}(1)$ ,  $C_2 = \mathcal{O}(1)$ , and a critical regime where  $C_1 = \mathcal{O}(1/\sqrt{N})$ ,  $C_2 = \mathcal{O}(1/N)$ . Scale bars on the  $x$  and  $y$  axis represent one unit of  $\ln C_1$  and  $\ln C_2$  respectively. In the high temperature phase and in the critical regime,  $\Lambda$  is a function of the scaled correlations  $NC_1$ ,  $N^2C_2$  and  $\sqrt{N}C_1$ ,  $NC_2$  respectively.

and

$$\mathcal{M} = \begin{pmatrix} x^2(\lambda_1 + 2\lambda_2 C_I) & x(1-x)(\lambda_1 + 2\lambda_2 C_{II}) \\ x(1-x)(\lambda_1 + 2\lambda_2 C_{II}) & (1-x)^2(\lambda_1 + 2\lambda_2 C_I) \end{pmatrix}. \quad (12)$$

Thus, at fixed  $x$  we can solve Eqs. (9,10), and then we can find the value of  $x$  that maximizes  $s$ . Along the same lines, one can solve the maximum entropy problem in the low temperature phase where  $C_1 = \mathcal{O}(1)$ ,  $C_2 = \mathcal{O}(1)$ . We can thus draw a phase diagram directly in the space of the observed moments  $C_1$  and  $C_2$ , which we plot as follows: we represent the partition function as an integral over two order parameters  $(m_A, m_B) \equiv \vec{m}$ , i.e.  $Z \propto \int d\vec{m} \exp[-N\mathcal{S}(\vec{m})]$ , where

$$\begin{aligned} \mathcal{S}(\vec{m}) &= \frac{1}{2} \vec{m}^T \cdot \mathcal{M}^{-1} \cdot \vec{m} - x \log 2 \cosh(\sqrt{2}m_A/x) + \\ &\quad - (1-x) \log 2 \cosh(\sqrt{2}m_B/(1-x)). \end{aligned} \quad (13)$$

Then, we consider the smallest Hessian eigenvalue  $\Lambda$  of  $\mathcal{S}$  computed at the saddle point  $\vec{m}^*$  of the integral in  $Z$ , and in Fig. 2 we show contour plots of  $\Lambda$  evaluated at the solution of the maximum entropy problem. There is a high temperature phase  $C_1 = \mathcal{O}(1/N)$ ,  $C_2 = \mathcal{O}(1/N^2)$ , and a low temperature phase  $C_1 = \mathcal{O}(1)$ ,  $C_2 = \mathcal{O}(1)$ : in both these phases  $\Lambda = \mathcal{O}(1)$ . The high and low temperature phases are separated by a critical regime  $C_1 = \mathcal{O}(1/\sqrt{N})$ ,  $C_2 = \mathcal{O}(1/N)$  where  $\Lambda = \mathcal{O}(1/\sqrt{N})$ . We see that  $\Lambda$  gets small close to the boundary of the allowed values  $C_2 = C_1^2$ : the contours of constant  $\Lambda$  bend backward, however, suggesting that we can reach a critical regime  $\Lambda = \mathcal{O}(1/\sqrt{N})$  if  $C_2 = \mathcal{O}(1/N)$  and  $C_1 = \mathcal{O}(1/N)$ , and we have verified this analytically.

How does Fig. 2 relate to what we know about the phase diagrams of Ising models? For a ferromagnetic Ising model on a  $d$ -dimensional hypercube with nearest-neighbor couplings  $J_{ij} = J > 0$ , the correlations at the critical point are  $C_1 \sim \sqrt{C_2} \sim 1/N^\omega$ , with  $\omega = (d - 2 + \eta)/d$ ; we have  $\omega = 1/8, 0.3455 \dots, 1/2$  for  $d = 2, 3, 4$ , respectively. Keeping just two moments our maximum entropy model matches the mean field critical behavior expected at  $d = 4$ . For an Ising model on a  $d$ -dimensional hypercube with antiferromagnetic nearest-neighbor interactions  $J_{ij} = J < 0$ , the lattice can be divided into two embedded sublattices with spins up and down respectively: as a result, roughly half of the spin pairs are positively correlated and the other half are negatively correlated, so  $C_1 \sim 1/N$ , and in the critical regime  $C_2 \sim 1/N^{2\omega} \gg C_1^2$ : This is again in reasonably good agreement with the model prediction that there is a critical behavior for  $C_1 = \mathcal{O}(1/N)$ ,  $C_2 = \mathcal{O}(1/N)$ . We note that both in the one moment and two moment case our maximum entropy solution provides a mean field critical scaling  $C_1 \sim \sqrt{C_2} \sim 1/N^{1/2}$ , where the fractional exponent  $1/2$  results from a Taylor expansion of the free energy in the neighborhood of the high-temperature saddle point  $\vec{m}^* = \vec{0}$ . In order to obtain a non-mean-field exponent, one needs to consider an infinite number of moments, so that the free energy can become a non-analytic function of  $\vec{m}$  at the critical point. To conclude, the analysis for the case where we fix just two moments in the distribution of pairwise correlations seems (barely) sufficient to identify critical behavior in simple ferromagnets and antiferromagnets. In addition, by fixing at least three moments of the correlations, one obtains a maximum entropy solution with frustrated spin-spin interactions which describes a spin glass (see Supplemental Material).

*Discussion* – The maximum entropy method is an approach to build models for the joint probability distribution of many degrees of freedom, making use only of measured expectation values for a limited number of operators. This approach is attractive both because it picks out the least structured model consistent with the data and because this model is exactly a statistical mechanics problem—the Boltzmann distribution with an energy landscape composed of a sum of terms, one for each of the measured expectation values. As noted at the outset, the maximum entropy construction can thus be thought of as an inverse statistical mechanics problem, mapping expectation values back to the coupling constants in the effective Hamiltonian. In this work we have used the maximum entropy method, but in looking at a strongly inhomogeneous system we have constrained not the expectation values of each local operator, but rather the distribution of these expectation values across the (large) system. In this sense, what we have constructed is an inverse spin glass: rather than choosing coupling constants at random from a distribution and calculating correlations, we choose the correlations from a distribution and infer the coupling constants.

In the Sherrington–Kirkpatrick spin glass, complete randomness in the coupling constants drives the emergence a rich, ordered structure in the pattern of correla-

tions among spins [28]. Here we have seen that knowing only the distribution of correlations leads to surprising structure in the coupling constants: the values of the spin couplings break into blocks, and with this ansatz we can derive a phase diagram in the space of moments. Although this needs to be checked carefully, it is natural to conjecture that the number of blocks grows with the number of moments that we constrain.

We thank A Cavagna, P Del Giudice, I Giardina, E Marinari, T Mora, G Parisi, and G Tkačik for helpful discussions. WB is especially grateful to his colleagues in Rome for their hospitality on several visits that influenced this work. Research supported in part by NSF Grants PHY-0957573, PHY-1305525 and CCF-0939370, by the Human Frontiers Science Program, by the Swartz Foundation, and by the WM Keck Foundation.

- 
- [1] ET Jaynes, Information theory and statistical mechanics. *Phys Rev* **106**, 620–630 (1957).
  - [2] E Schneidman, MJ Berry II, R Segev, and W Bialek, *Nature* **440**, 1007–1012 (2006); arXiv:q-bio.NC/0512013 (2005).
  - [3] J Shlens, GD Field, JL Gauthier, MI Grivich, D Petrusca, A Sher, AM Litke, and EJ Chichilnisky, *J Neurosci* **26**, 8254–8266 (2006).
  - [4] G Tkačik, E Schneidman, MJ Berry II, and W Bialek, arXiv:q-bio/0611072 (2006).
  - [5] S Yu, D Huang, W Singer, and D Nikolic, *Cereb Cortex* **18**, 2891–2901 (2008).
  - [6] A Tang *et al.*, *J Neurosci* **28**, 505–518 (2008).
  - [7] G Tkačik, E Schneidman, MJ Berry II, and W Bialek, arXiv:0912.5409 (2009).
  - [8] J Shlens, GD Field, JL Gauthier, M Greschner, A Sher, AM Litke, and EJ Chichilnisky, *J Neurosci* **29**, 5022–5031 (2009).
  - [9] IE Ohiorhenuan, F Mechler, KP Purpura, AM Schmid, Q Hu, and JD Victor, *Nature* **466**, 617–621 (2010).
  - [10] E Gannor, R Segev, and E Schneidman, *Proc Natl Acad Sci (USA)* **108**, 9679–9684 (2011).
  - [11] G Tkačik, O Marre, T Mora, D Amodei, MJ Berry II, and W Bialek, *J Stat Mech* P03011 (2013); arXiv:org:1207.6319 (2012).
  - [12] TR Lezon, JR Banavar, M Cieplak, A Maritan, and NV Fedoroff, *Proc Natl Acad Sci (USA)* **103**, 19033–19038 (2006).
  - [13] G Tkačik, *Information Flow in Biological Networks* (Dissertation, Princeton University, 2007).
  - [14] W Bialek and R Ranganathan, arXiv:0712.4397 [q-bio.QM] (2007).
  - [15] F Seno, A Trovato, JR Banavar, and A Maritan, *Phys Rev Lett* **100**, 078102 (2008).
  - [16] M Weigt, RA White, H Szurmant, JA Hoch, and T Hwa, *Proc Natl Acad Sci (USA)* **106**, 67–72 (2009).
  - [17] N Halabi, O Rivoire, S Leibler, and R Ranganathan, *Cell* **138**, 774–786 (2009).
  - [18] T Mora, AM Walczak, W Bialek, and CG Callan Jr, *Proc Natl Acad Sci (USA)* **107**, 5405–5410 (2010); arXiv:0912.5175 [q-bio.GN] (2009).
  - [19] DS Marks, LJ Colwell, R Sheridan, TA Hopf, A Pagnani, R Zecchina, and C Sander, *PLoS One* **6**, e28766 (2011).
  - [20] JI Sulkowska, F Morcos, M Weigt, T Hwa, and JN Onuchic, Genomics-aided structure prediction. *Proc Natl Acad Sci (USA)* **109**, 10340–10345 (2012).
  - [21] W Bialek, A Cavagna, I Giardina, T Mora, E Silvestri, M

- Viale, and A Walczak, *Proc Natl Acad Sci (USA)* **109**, 4786–4791 (2012); arXiv.org:1107.0604 [physics.bio-ph] (2011).
- [22] W Bialek, A Cavagna, I Giardina, T Mora, O Pohl, E Silvestri, M Viale, and A Walczak, arXiv.org:1307.5563 [physics.bio-ph] (2013).
- [23] S Cocco, R Monasson, and V Sessak, *Phys Rev E* **83**, 051123 (2011).
- [24] F Ricci–Tersenghi, *J Stat Mech* P08015 (2012).
- [25] A Decelle, F Ricci–Tersenghi, *Phys Rev Lett* **112**, 070603 (2014).
- [26] J Kaipio and E Somersalo, *Statistical and Computational Inverse Problems* (Springer, New York, 2005).
- [27] S Cocco and R Monasson, *Phys Rev Lett* **106**, 090601 (2011).
- [28] M Mézard, G Parisi, and MA Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

# Supplemental Material for “Inverse spin glass and related maximum entropy problems”

Michele Castellana<sup>1</sup> and William Bialek<sup>1,2</sup>

<sup>1</sup>*Joseph Henry Laboratories of Physics and Lewis–Sigler Institute for Integrative Genomics,  
Princeton University, Princeton, New Jersey 08544*

<sup>2</sup>*Initiative for the Theoretical Sciences, The Graduate Center,  
City University of New York, 365 Fifth Ave., New York, New York 10016*

(Dated: September 7, 2014)

## INDEPENDENT SPINS

Here we show how to obtain the maximum entropy solution shown in Fig. 1 where we constrain all moments of the magnetization in the independent spin case. The Lagrange function (2) reads

$$\begin{aligned} \mathcal{L} = & - \sum_{\sigma} P(\sigma) \ln P(\sigma) + N \lambda_0 \left[ \sum_{\sigma} P(\sigma) - 1 \right] + (S1) \\ & + N \int dm \lambda(m) \left[ \frac{1}{N} \sum_{i=1}^N \delta(\langle \sigma_i \rangle - m) - p(m) \right], \end{aligned}$$

where  $\{\lambda(m)\}_m$  is a continuum set of Lagrange multipliers enforcing the constraint of the magnetization distribution  $p(m)$  for every  $m$ . By taking  $\partial \mathcal{L} / \partial P(\sigma) = 0$  we have

$$P(\sigma) = \frac{1}{Z} \exp \left( \sum_{i=1}^N h_i \sigma_i \right), \quad (S2)$$

where

$$h_i = - \left. \frac{d \lambda(m)}{dm} \right|_{m=\langle \sigma_i \rangle}. \quad (S3)$$

The magnetic fields  $h_i$  are related to the local magnetizations  $\langle \sigma_i \rangle$  by the equation

$$\langle \sigma_i \rangle = \tanh h_i. \quad (S4)$$

From Eq. (S4) we can compute the distribution of fields

$$\begin{aligned} \mathfrak{p}_{\text{all}}(h) &= \frac{1}{N} \sum_{i=1}^N \delta(h_i - h) \quad (S5) \\ &= p(\tanh(h)) [1 - \tanh(h)^2]. \end{aligned}$$

The entropy per spin is given by

$$s = - \frac{1}{N} \sum_{\sigma} P(\sigma) \ln P(\sigma), \quad (S6)$$

and from Eq. (S5) we obtain the entropy in the all-moment case

$$s_{\text{all}} = \int dh \mathfrak{p}_{\text{all}}(h) [\ln(2 \cosh h) - h \tanh h]. \quad (S7)$$

If we constrain only the first two moments, it can be shown that the magnetic fields  $h_i$  can take only two values: indeed, let us fix  $\lambda_1, \lambda_2$  such that there are three solutions for  $\langle \sigma_i \rangle$  to Eq. (5), that we call  $m_A, m_B, m_C$ . Let us denote by  $x_A, x_B, x_C$ , with  $x_B = 1 - (x_A + x_C)$ , the fraction of spins with  $\langle \sigma_i \rangle$  equal to  $m_A, m_B, m_C$  respectively: we determine  $x_A, x_C$  by solving the two moment constraint

$$\begin{aligned} m_1 &= x_A m_A + (1 - (x_A + x_C)) m_B + x_C m_C, \\ m_2 &= x_A m_A^2 + (1 - (x_A + x_C)) m_B^2 + x_C m_C^2, \end{aligned}$$

and the entropy  $s$  now depends only on  $\lambda_1, \lambda_2$ . If we now maximize  $s$  with respect to  $\lambda_1, \lambda_2$ , the maximal entropy is obtained for  $x_C = 0$ , and it coincides with the one obtained with our two block ansatz where the magnetic fields take only two values  $h_A, h_B$ . Along the same lines, if we constrain more than two moments, the correct number of distinct solutions  $h_A, h_B, \dots$  of the analog of Eq. (5) can be singled out as the solutions providing the largest entropy.

## COUPLED SPINS

### One moment case

Here we will discuss briefly the maximum entropy solution where only the first correlation moment is constrained: more details will be given in the calculation for the two moment case. If we constrain only  $C_1$ , then Eq. (7) becomes the familiar mean-field ferromagnet: the partition function  $Z$  can be written as an integral over the order parameter, i.e.  $Z \propto \int dm \exp[-N \mathcal{S}(m)]$ . So long as  $NC_1$  is of order unity the system is in its high temperature phase. If  $NC_1$  is of order  $\sqrt{N}$ , however,  $\lambda_1$  is driven close enough to its critical value that we must go to higher order in the expansion around the saddle point  $m^*$  of the integral in  $Z$ . When  $\sqrt{N} C_1 > 1.17083 \dots$ , there is no consistent solution in the high temperature phase: the only way to match the observed  $C_1$  is if this mean correlation has a contribution from a nonzero mean magnetization, and the system is in its low temperature phase. In order to draw the phase diagram of the system, we consider the second derivative of  $\mathcal{S}(m)$  at the saddle point  $\Lambda \equiv d^2 \mathcal{S} / dm^2|_{m^*}$ . This quantity indicates the stability of the saddle point: a small value of  $\Lambda$  implies that the saddle point is nearly unstable, and thus the system

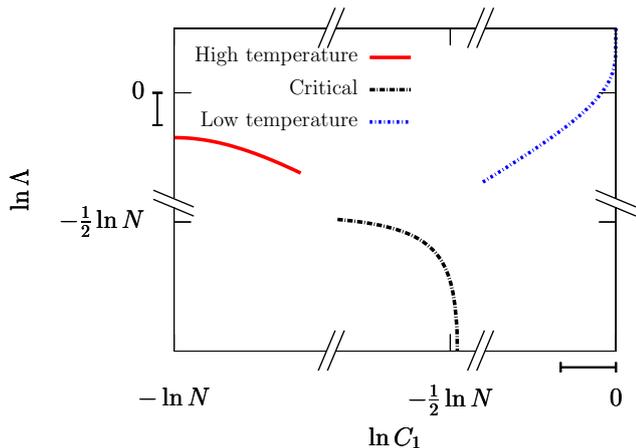


FIG. S1: Phase diagram for coupled spins in the one moment case: second derivative  $\Lambda$  of  $\mathcal{S}(m)$  at the saddle point  $m = m^*$  as a function of  $C_1$  in the high temperature phase where  $C_1 = \mathcal{O}(1/N)$  and  $\Lambda = \mathcal{O}(1)$  (in red), in the critical region where  $C_1 = \mathcal{O}(1/\sqrt{N})$  and  $\Lambda = \mathcal{O}(1/\sqrt{N})$  (in black), and in the low temperature phase where  $C_1 = \mathcal{O}(1)$  and  $\Lambda = \mathcal{O}(1)$  (in blue). The scale bars on the  $x$  and  $y$  axes represent one unit of  $\ln C_1$  and  $\ln \Lambda$  respectively.

is close to a critical point. In Fig. S1 we plot  $\Lambda$  vs.  $C_1$  in the high and low temperature phase and in the critical region: thus, even in the problem where we constrain only one moment of the distribution of correlations, we can draw a phase diagram along the  $C_1$  axis, without reference to any other parameters.

### Two moment case

First, we performed numerical experiments to verify that Eq. (8) has only two solutions: we fixed  $\lambda_1, \lambda_2$  and we solved Eq. (8) for the variables  $\{J_{ij}\}$ . The solution of these equations involves the exact computation of the correlation functions  $\langle \sigma_i \sigma_j \rangle$ , which involves  $2^N$  terms: hence, this numerical experiment is limited to small values of  $N$ . An example of this computation is shown in Fig S2: for  $N = 4$  the distribution of coupling is given by three peaks, while as  $N$  is increased the distribution is driven to consist of two peaks.

Let us now discuss the solution of the maximum entropy problem within the two block ansatz: the magnetization in block A is

$$m_A = \frac{1}{N_A} \sum_{i=1}^{N_A} \sigma_i, \quad (\text{S8})$$

and similarly for  $m_B$ . By using the block ansatz for the correlation matrix  $C_{ij}$  shown in Fig. S3, we can write Eq. (7) as

$$P(\sigma) = \frac{1}{Z} \exp(N \vec{m}^T \cdot \mathcal{M} \cdot \vec{m}), \quad (\text{S9})$$

where  $\vec{m} \equiv (m_A, m_B)$  and  $\mathcal{M}$  is given by Eq. (12). The sum over spin configurations can be written as an integral over the order parameters  $m_A$  and  $m_B$  by standard

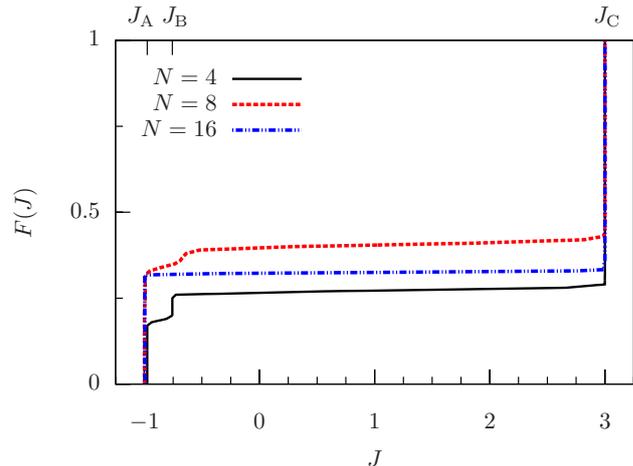


FIG. S2: Cumulative distribution function of couplings for coupled spins in the two moment case. For a given  $N$  and  $\lambda_1 = \lambda_2 = 1$ , Eq. (8) is solved numerically multiple times starting from different random initial conditions for  $\{J_{ij}\}$ : as a result, a total number  $512 \leq S \leq 65536$  of couplings  $J_{ij}$  is obtained from this ensemble of solutions. The cumulative distribution function  $F(J)$  is defined as the number of couplings  $J_{ij}$  smaller than  $J$  divided by the total number of couplings  $S$ , and it is plotted as a function of  $J$  for system sizes  $N = 4, 8, 16$ . For  $N = 4$ , the  $J_{ij}$ s cluster tightly around three values  $J_A, J_B, J_C$ , and by  $N = 16$  the distribution of  $J_{ij}$ s is almost perfectly discrete, with only two values  $J_A, J_C$ .

$$C_{ij} = \left( \begin{array}{cc} C_I & C_{II} \\ C_{II} & C_I \end{array} \right) \left. \begin{array}{l} \vphantom{\left( \right)} \\ \vphantom{\left( \right)} \end{array} \right\} \begin{array}{l} N_A \\ N_B \end{array}$$

$\underbrace{\hspace{10em}}_N$

FIG. S3: Block ansatz for the correlation matrix  $C_{ij}$ : correlations between spins within a block are equal to  $C_I$ , while correlations between spins in different blocks are equal to  $C_{II}$ .

methods. The result for the partition function is

$$Z = \frac{2\pi N}{\sqrt{\det \mathcal{M}}} \int d\vec{m} \exp[-N\mathcal{S}(\vec{m})], \quad (\text{S10})$$

where  $\mathcal{S}$  is given by Eq. (13). The integral in Eq. (S10) is dominated by the saddle point  $\vec{m}^*$  which minimizes  $\mathcal{S}$ : let us consider first the high temperature case, where we match the observed correlations with a model with saddle point  $\vec{m}^* = \vec{0}$ . The free energy is

$$f = -\ln 2 + \frac{1}{2N} \ln \mathcal{D}, \quad (\text{S11})$$

where  $\mathcal{D} = \det(1 - 2 \text{diag}(x, 1-x)^{-1} \cdot \mathcal{M})$ . The entropy per spin is given by Eq. (S6), and Eq. (11) is then obtained from Eqs. (S6, S11). The entropy (S6) can be written as the difference between an energy term and the free energy. Hence, the maximum of the entropy is obtained as a tradeoff between free energy and energy.

This maximum is reminiscent of the minimum of the free energy in the direct problem, which is obtained as a tradeoff between energy and entropy.

In the low temperature case, the saddle point is at  $\vec{m} = \vec{m}^* \neq \vec{0}$ . The free energy is given by  $f = \mathcal{S}(\vec{m}^*)$ , and it can be computed numerically for any given  $\lambda_1, \lambda_2, x$ : the optimal value of  $x$  can be obtained along the same lines as in the high temperature case.

### Three moment case

Here we will show that our maximum entropy solution reproduces an Ising spin glass if we constrain three moments of the correlations. The main feature of an Ising spin glass is the presence of frustration: namely, the product of the interactions  $J_{ij}$  along a loop of spins has a negative sign. In order to obtain a frustrated maximum entropy solution, one needs to split the coupling matrix  $J_{ij}$  into at least three blocks. As a consequence, a spin-glass maximum entropy solution can be obtained only if

at least three moments of the correlations are constrained. This can be done along the lines of the two moment case, and we will sketch the main steps here. We maximize the entropy (S6) by constraining the first three moments  $C_1, C_2, C_3$  of the correlations: the Boltzmann distribution is still given by Eq. (7), with

$$J_{ij} = \lambda_1 + 2\lambda_2 C_{ij} + 3\lambda_3 C_{ij}^2. \quad (\text{S12})$$

We construct a three block correlation matrix  $C_{ij}$  starting from the two block case: we consider the bottom-right block in Fig. S3, and we partition it into blocks as we did for the matrix  $C_{ij}$  in the two moment case. As a result, spins are now divided into groups  $A \equiv \{\sigma_1, \dots, \sigma_{N_A}\}$ ,  $B \equiv \{\sigma_{N_A+1}, \dots, \sigma_{N_A+N_B}\}$  and  $C \equiv \{\sigma_{N_A+N_B+1}, \dots, \sigma_N\}$ , and correlations  $C_{ij}$  take three values  $C_I, C_{II}, C_{III}$ . In the high temperature phase where  $C_1 = \mathcal{O}(1/N)$ ,  $C_2 = \mathcal{O}(1/N^2)$ ,  $C_3 = \mathcal{O}(1/N^3)$ , the free energy is given by Eq. (S11), with  $\mathcal{D} = \det(1 - 2 \text{diag}(x, y, 1 - x - y)^{-1} \cdot \mathcal{M})$  where  $x = N_A/N$ ,  $y = N_B/N$ , and the matrix  $\mathcal{M}$  is given by

$$\mathcal{M} = \begin{pmatrix} x^2(\lambda_1 + 2\lambda_2 C_I + 3\lambda_3 C_I^2) & x(y\lambda_1 + 2u\lambda_2 C_{II} + 3u\lambda_3 C_{II}^2) & x(z\lambda_1 + 2u\lambda_2 C_{II} + 3u\lambda_3 C_{II}^2) \\ x(y\lambda_1 + 2u\lambda_2 C_{II} + 3u\lambda_3 C_{II}^2) & y^2(\lambda_1 + 2\lambda_2 C_I + 3\lambda_3 C_I^2) & yz(\lambda_1 + 2\lambda_2 C_{III} + 3\lambda_3 C_{III}^2) \\ x(z\lambda_1 + 2u\lambda_2 C_{II} + 3u\lambda_3 C_{II}^2) & yz(\lambda_1 + 2\lambda_2 C_{III} + 3\lambda_3 C_{III}^2) & z^2(\lambda_1 + 2\lambda_2 C_I + 3\lambda_3 C_I^2) \end{pmatrix}, \quad (\text{S13})$$

with  $z = 1 - x - y$ ,  $u = 1 - x$ . Given  $C_1, C_2, C_3$  and  $x, y$ , we determine  $C_I, C_{II}, C_{III}$  and  $\lambda_1, \lambda_2, \lambda_3$  by matching the moments of correlations as in Eq. (9), and then we maximize  $f$  with respect to  $x, y$ .

In an Ising spin glass we expect spin-spin correlations to have both positive and negative signs: hence, the odd moments of correlations are small compared to the even ones. Indeed, if the correlation moments satisfy these conditions, we obtain that our maximum entropy solution involves frustrated spin-spin interactions typical of a spin glass. Explicitly, we take  $C_1 = 0.1$ ,  $C_2 = 0.5$ ,  $C_3 = 0.05$ , and we obtain that the interactions along a loop connecting spins in blocks A, B, C are  $J_{AB} = -0.13$ ,  $J_{BC} = -0.098$ ,  $J_{CA} = J_{AB}$ , implying that  $J_{AB}J_{BC}J_{CA} < 0$ .

As we fix more and more moments, we expect our maximum entropy method to reproduce the full features of a spin glass, such as the emergence of a critical behavior. To identify the scaling of correlations cor-

responding to the critical regime, we recall that given the overlap  $Q \equiv \frac{1}{N} \sum_{i=1}^N \sigma_i^1 \sigma_i^2$  between two independent replicas  $\sigma^1, \sigma^2$ , in the high temperature regime we have  $\langle Q^2 \rangle = \mathcal{O}(1/N)$  [1], hence for large  $N$

$$\begin{aligned} C_2 &= \frac{1}{N^p} \sum_{i>j} \langle \sigma_i \sigma_j \rangle^2 \\ &= \frac{1}{(1 - 1/N)} \left( \frac{1}{N^2} \sum_{ij} \langle \sigma_i \sigma_j \rangle^2 \right) - \frac{1}{N} \\ &= \langle Q^2 \rangle - \frac{1}{N}. \end{aligned} \quad (\text{S14})$$

It follows that the critical spin glass regime is obtained as  $C_1 \approx C_3 \approx 0$ ,  $C_2 \gg \mathcal{O}(1/N)$ .

[1] M Mézard, G Parisi, and MA Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).