Number of Stable Points for Spin-Glasses and Neural Networks of Higher Orders

Pierre Baldi[a]
Department of Mathematics, California Institute of Technology, Pasadena, California 91125
and
Santosh S. Venkatesh[b]
Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125
(Received 23 June 1986)

We study the number of stable points for spin-glasses and neural networks of higher orders, i.e., with Hamiltonians given by an algebraic form of degree d. For spin-glasses, we derive a rigorous exact expression, in the thermodynamic limit, assuming long-range independent exchange Gaussian interactions among sets of d spins. For neural networks we introduce several upper bounds on the number of programmable stable states, according to different storage schemes.

PACS numbers: 75.10.Hk

In recent years, considerable progress has been made in the study of disordered systems like spin-glasses. In addition, several models[1,2] for large interconnected networks of neurons with emergent collective behavior have been proposed by use of ideas borrowed from statistical mechanics. Applications ranging from content-addressable memories and new integrated circuit designs to optimization and learning algorithms[3] have been partially investigated. The dynamics of spin-glasses can be used to perform computations such as error correction and nearest-neighbor search, and information can be stored by sculpting of the energy valleys of these systems. Along the same lines, interesting connections between intractable combinatorial problems and statistical mechanics based on the landscape of the free-energy function[4] have resulted in new optimization algorithms (simulated annealing).

With the assumption of a system of N Ising spins $x_i = \pm 1$, most of these studies are based on a classical quadratic Hamiltonian of the form

$$H(x) = -\tfrac{1}{2}\sum_{i,j} T_{i,j}x_i x_j.$$

Neural networks of this sort, however, have several limitations. As an example, the number of prescribed fixed points that can be realized in such systems is of the order of N. To achieve greater flexibility and programming capability several authors[5-7] have noticed that Hamiltonians of higher order, i.e., defined by an algebraic form of degree d, could easily be introduced. Moreover, such forms arise naturally in optimization problems, for instance, and can be implemented in optical and electronic circuits in a fairly simple fashion.

In the following, we first precisely define such higher-order systems. We then analyze the number of stable states in two very distinct situations. We start with the long-range spin-glass case at zero temperature, where the coefficients of the Hamiltonian are assumed to be random independent, identically distributed (i.i.d.) Gaussian variables. We prove a precise general formula which gives the number of such points in the thermodynamic limit ($N \to \infty$). We proceed with the neural-network case, where the coefficients are carefully constructed according to a certain storage rule in order to force a prechosen set of vectors to be among the stable states of the network. We state upper bounds on the storage capacity of these systems under different storing strategies. Detailed mathematical proofs will be given elsewhere.

Let $\mathcal{I}$ be the family of all subsets of cardinal d of the set $\{1,2,\ldots,N\}$. Therefore, $|\mathcal{I}| = \binom{N}{d}$. For any subset J of $\{1,2,\ldots,N\}$, let $x^J = \prod_{j\in J} x_j$. We can consider the system described (up to a multiplicative constant) by the homogeneous Hamiltonian of degree d,

$$H(x) = \sum_{I\in\mathcal{I}} T_I x^I, \tag{1}$$

where the T_I are real coefficients indexed by I in $\mathcal{I}$. Factoring x_i we can write

$$H(x) = x_i\left(\sum_{i\in I} T_I x^{I-\{i\}}\right) + \sum_{i\notin I} T_I x^I.$$

We start from an arbitrary configuration and let the system evolve by a sequence of single spin flips, involving spins which are misaligned with their instantaneous "molecular field." In other words, we have the evolution rule

$$x_i^+ = \operatorname{sgn}\left(\sum_{i\in I} T_I x^{I-\{i\}}\right). \tag{2}$$

This can be seen essentially as a zero-temperature Monte Carlo (or Glauber) dynamics. Because of the monotonicity of (1) under the rule (2) the system always reaches a stable state where the relation (2) is satisfied for each one of the N spins. This form of exchange interaction lends itself to natural extensions of spin-glass

formulations as well as to certain neural-network models. In the next two sections, we will consider the number of stable points from these two different perspectives.

Higher-order spin-glasses.— Let T_I be i.i.d. $\mathcal{N}(0,1)$, i.e., independent, identically distributed normal random variables with zero mean and variance 1. Let F_N^d be the expected number of stable states. The purpose is to estimate F_N^d, for N large. We generalize a result of Edwards and Tanaka[8] and improve on their precision with a technique used by McEliece and Posner[9] for the usual case $d=2$.

We can write $F_N^d = 2^N P_N^d$, where P_N^d is the probability that a given state x is stable. Without any loss of generality we can assume $x=1,1,\ldots,1$. Therefore, denoting probabilities by the symbol Pr, we have

$$P_N^d = \Pr(S_i > 0,\ i=1,2,\ldots,N)\ , \tag{3}$$

where $S_i = \sum_{i\in I} T_I$. The N variables S_i are $\mathcal{N}(0, \binom{N-1}{d-1})$ and $E(S_iS_j) = \binom{N-2}{d-2}$, for $i\neq j$. We use the method of equivalent Gaussians[10] to estimate the number of fixed points. Let $Y_0,\ldots,Y_N$ be $N+1$ i.i.d. $\mathcal{N}(0,1)$. Construct N new Gaussian variables V_i with the same properties as the variables S_i by letting

$$V_i = \alpha_N Y_i + \beta_N Y_0,\quad i=1,\ldots,N, \tag{4}$$

with $\alpha_N = \binom{N-2}{d-1}^{1/2}$, and $\beta_N = -\binom{N-2}{d-2}^{1/2}$. We then obtain $F_N^d = 2^N \Pr(V_i > 0,\ 1\le i\le N)$, and so $F_N^d = 2^N \Pr[Y_i > -(\beta_N/\alpha_N)Y_0,\ \ 1\le i\le N]$. The variables $Y_1,\ldots,Y_N$ are conditionally independent given that $Y_0 = y_0$. Hence,

$$F_N^d = 2^N \int_{-\infty}^{+\infty} \frac{1}{(2\pi)^{1/2}} e^{-t^2/2} \left[\Phi\left(\frac{\beta_N}{\alpha_N} t\right)\right]^N dt, \tag{5}$$

where

$$\Phi(z) = (2\pi)^{-1/2} \int_{-\infty}^{z} e^{-t^2/2} dt$$

is the cumulative Gaussian distribution function. Now $-\beta_N/\alpha_N = [(d-1)/(N-d)]^{1/2}$. If we make the change of variable $u = -(\beta_N/\alpha_N)t$, we get

$$F_N^d = 2^N \left(\frac{N-d}{d-1}\right)^{1/2} \int_{-\infty}^{+\infty} \{\Phi(u)\exp[-u^2/2(d-1)]\}^N \exp[du^2/2(d-1)]du. \tag{6}$$

The integral in (6) can be estimated by use of Laplace or saddle-point methods.[11] If we let $e^{h(u)} = \Phi(u) \times \exp[-u^2/2(d-1)]$ and $g(u) = \exp[du^2/2(d-1)]$ then, as N goes to infinity,

$$\int_{-\infty}^{+\infty} e^{Nh(u)} g(u)du \simeq \frac{e^{Nh(u_0)}}{N^{1/2}} g(u_0) \left(\frac{2\pi}{-h''(u_0)}\right)^{1/2}, \tag{7}$$

with $u_0 = \arg\max h(u)$. Details of proofs and calculations will be published.[12]

The general result is that the expected number of fixed points is given by an expression of the form

$$F_N^d \simeq k_d 2^{c_d N}, \tag{8}$$

where k_d and c_d are constants depending purely on the interaction order d and which can be precisely computed.

It should be noticed that without any loss of generality we can assume $d \le \frac{1}{2}N$. When d is fixed, we obtain

$$k_d = \exp\left(\frac{du_0^2}{2d-2}\right)\left(\frac{d-1}{du_0^2+d-1}\right)^{1/2}, \tag{9}$$

$$c_d = 0.5 + \log_2\left(\frac{d-1}{\pi^{1/2}|u_0|}\right) - \frac{u_0^2 d}{2(d-1)\ln 2}, \tag{10}$$

where u_0 is the unique solution of the equation

$$e^{-u_0^2/2}/(2\pi)^{1/2}\Phi(u_0) - u_0/(d-1) = 0. \tag{11}$$

If d scales with N such that $d \le \frac{1}{2}N$ and $d \to +\infty$, then

$$k_d = d/2(2\pi)^{1/2}\ln d, \tag{12}$$

$$c_d = 1 - \ln\ln d/(2\ln 2)\ln d. \tag{13}$$

If $d/N = \text{const} = K$, then

$$k_d = KN/2(2\pi)^{1/2}\ln N, \tag{14}$$

$$c_d = 1 - \ln\ln N/(2\ln 2)\ln N. \tag{15}$$

Table I indicates the behavior of F_N^d for some fixed values of d.

Higher-order neural networks.— In the case of neural networks the problem is rather different. With the assumption that k vectors $M^1,\ldots,M^k$ of dimension N with ± 1 coordinates have been selected, the purpose is

TABLE I. Indication of the behavior of $F_N^d \simeq k_d 2^{c_d N}$ for different fixed values of the degree d.

d	k_d	c_d
1	1	0
2	1.0505	0.2874
3	1.1320	0.4265
4	1.2178	0.5124
5	1.3031	0.5721
10	1.7032	0.7215
50	4.1667	0.9090
100	6.6705	0.9461
1000	39.3100	0.9916

to construct a nontrivial energy function of degree d having these vectors among its local minima over the hypercube. The dynamical evolution of the network can then be interpreted in terms of computations: error correction, associative memory, and nearest-neighbor search. Here we are not interested, in the first place, in the total number of stable states but rather in the maximal possible value of k, which is often regarded as a measure for the storage capacity of a network. However, this capacity depends on the storage rule adopted, on the configuration and mode of selection of the vectors (orthogonal, at random, . . .), and on the quality required on the retrieval (are small errors acceptable or not?). In a probabilistic context, the maximal possible rate of growth of k with N has been used as an indicator of capacity.[7]

Various storage rules are reviewed by Baldi.[13] We shall here state results for three different cases: arbitrary rules, generalized sum of outer products rule, and generalized spectral rule. Detailed proofs, which are rather technical, are given in Ref. 12. Complements and partial results (especially for the quadratic case) can be found in the work of Venkatesh,[7] McEliece *et al.*,[14] Venkatesh,[15] and Venkatesh and Psaltis.[16]

(1) Arbitrary rules: If k is the largest integer such that for almost any set of k vectors $M^1, \ldots, M^k$, we can find a nontrivial form of degree d having $M^1, \ldots, M^k$ among its stable points, then

$$(N+1)^{-1} \sum_{j=0}^{d} \binom{N}{j} \leq k \leq \sum_{j=0}^{d-1} \binom{N-1}{j}.$$

(2) Generalized sum of outer products rule: This is a natural extension of the classical Hebbian rule for $d=2$. The coefficients T_I are constructed as the sum of generalized outer products,

$$T_I = \sum_{s=1}^{k} (M^s)^I.$$

If the components of the k vectors are chosen from a sequence of symmetric Bernouilli trials [i.e., $\Pr(M_i^s = -1) = \Pr(M_i^s = 1) = 0.5$], then the largest allowable rate of growth of k with N such that the k vectors are stable with probability approaching 1 as $N \to \infty$ is given by

$$k = N^{d-1}/2d! \ln N.$$

(3) Generalized spectral rule: When $d=2$, and given a starting state X, the spectral rule or projection rule amounts to iteratively projecting X orthogonally onto the linear space generated by $M^1, \ldots, M^k$, and then taking the closest point on the hypercube to this projection. This scheme can be extended to higher orders by use of pseudo inverse matrices. The algebraic construction is described in Ref. 12. Again the largest allowable rate of growth of k with N is

$$k = \sum_{j=1}^{d-1} \binom{N}{j}.$$

It should be remarked that in spite of differences in the assumptions, all these theorems point in a same direction: *For neural networks with energy function of fixed degree d, the maximal number of programmable stable states is essentially of the order of* $O(N^{d-1})$.

We have found precise estimates for the number of stable points of higher-order spin-glasses with infinite-range generalized Gaussian interactions. While the general exponential form of the solution can easily be guessed, the exact value of the parameters and the rate at which the number of local minima approaches 2^N, as the disorder increases with the order of the interactions, is nontrivial. We have described how certain models of neural networks and the corresponding storage rules can be generalized to include nonquadratic interactions and have estimated the corresponding capacities under different sets of assumptions. It may be possible to overcome some of the limitations of traditional neural networks by use of networks of degree d, since all our results indicate an increased capacity which scales as N^{d-1}. The precise effects of variations of d on the size of the basins of attraction of the patterns stored and on the total number of stable states (including those which are not in the list of initial patterns) requires additional work. Finally, it should be noticed that computer simulations are in very good agreement with our asymptotic estimates even for small values of N ($N \geq 20$).

[a]Present address: Department of Mathematics, University of California, San Diego, La Jolla, CA 92093.

[b]Present address: Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104.

[1]J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[2]W. A. Little, Math. Biosci. **19**, 101 (1974).

[3]J. J. Hopfield and D. W. Tank, Biol. Cybern. **52**, 141 (1985).

[4]S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671–680 (1983).

[5]P. Baldi, Ph.D. thesis, California Institute of Technology, 1986 (unpublished).

[6]P. Baldi, "Neural Networks, Orientations of the Hypercube and Algebraic Threshold Functions" (to be published).

[7]S. S. Venkatesh, Ph.D. thesis, California Institute of Technology, 1986 (unpublished).

[8]S. F. Edwards and F. Tanaka, J. Phys. F **10**, 2769–2778 (1980).

[9]R. J. McEliece and E. C. Posner, Jet Propulsion Laboratory Telecommunications and Data Acquisition Progress Report, Vol. 42-83 (unpublished), pp. 209–215.

[10]J. R. Pierce and E. C. Posner, *Introduction to Communica-*

tion Science and Systems (Plenum, New York, 1980).

[11]N. G. De Bruijn, *Asymptotic Methods in Analysis* (Dover, New York, 1958).

[12]P. Baldi and S. S. Venkatesh, "Capacities and Fixed Points for Generalized Neural Networks and Spin Glasses" (to be published).

[13]P. Baldi, "On the Storage Rules and Stable Points of a Family of Automats" (to be published).

[14]R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, "The Capacity of the Hopfield Associative Memory" (to be published).

[15]S. S. Venkatesh, in Proceedings of the Conference on Neural Networks for Computing, Epsilon Capacity of Neural Networks, Snowbird, Utah, 1986 (to be published).

[16]S. S. Venkatesh and D. Psaltis, "Information storage and retrieval in two associative nets" (to be published).