From shrinking to percolation in an optimization model

2000 J. Phys. A: Math. Gen. 33 L53

(http://iopscience.iop.org/0305-4470/33/7/101)

J. Phys. A: Math. Gen. **33** (2000) L53–L59. Printed in the UK PII: S0305-4470(00)04635-7

LETTER TO THE EDITOR

From shrinking to percolation in an optimization model

J van Mourik†‖, K Y Michael Wong‡ and D Bollé§

† Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK
‡ Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, People's Republic of China
§ Instituut voor Theoretische Fysica, K U Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium

Received 26 May 1999, in final form 21 October 1999

Abstract. A model of noise reduction for signal processing and other optimization tasks is introduced. Each noise source puts a symmetric constraint on the space of the signal vector within a tolerance bound. When the number of noise sources increases, sequences of transitions take place, causing the solution space to vanish. We find that the transition from an extended solution space to a shrunk space is retarded because of the symmetry of the constraints, in contrast with the analogous problem of pattern storage. For low tolerance, the solution space vanishes by volume reduction, whereas for high tolerance, the vanishing becomes more and more like percolation.

During the past few years, the statistical mechanics of disordered systems has been frequently applied to understanding the macroscopic behaviour of many technologically useful problems, such as optimization (e.g. graph partitioning and travelling salesman) [1], learning in neural networks [2], error correcting codes [3], K-satisfiability [4] and the number partitioning problem [5]. A focus of this approach is the phase transitions in such systems, e.g. the glassy transition in optimization when the noise temperature of the simulated annealing process is reduced, the storage capacity in neural networks and the entropic transition in the K-satisfiability and number partitioning problem. Understanding these transitions is relevant to the design and algorithmic issues in their applications [6]. In turn, since the behaviour may be distinct from conventional disordered systems, the perspectives of statistical mechanics are widened.

In this paper we consider the phase transitions in an optimization problem, which is applicable to noise reduction (NR) techniques in signal processing, and other tasks. They have been used in a number of applications such as adaptive noise cancellation, echo cancellation, adaptive beamforming and more recently, blind separation of signals [7, 8]. While the formulation of the problem depends on the context, the following model is typical. There are N detectors picking up signals mixed with noise from p noise sources. The input from detector j is $x_j = a_j S + \sum_\mu \xi_j^\mu n_\mu$, where S is the signal, n_μ for $\mu = 1, \ldots, p$ is the noise from the μth noise source, and $n_\mu \ll S$. The a_j and ξ_j^μ are the contributions of the signal and the μth noise source to detector j. NR involves finding a linear combination of the inputs so that the noises are minimized while the signal is kept detectable. Thus, we search for an N-dimensional vector J_j such that the quantities $\sum_j \xi_j^\mu J_j$ are minimized, while $\sum_j a_j J_j$ remains a non-zero constant. To consider solutions with comparable power, we add the constraint $\sum_j J_j^2 = N$.

‖ Author to whom correspondence should be addressed.

0305-4470/00/070053+07$30.00

While adaptive algorithms for this objective exist [7], here we are interested in whether the noise can be intrinsically kept below a tolerance level after the steady state is reached, provided a converging algorithm is available.

When both p and N are large, we use a formulation with normalized parameters. Let h^μ be the local fields for the μth source defined by $h^\mu \equiv \sum_j \xi_j^\mu J_j/\sqrt{N}$. Learning involves finding a vector J_j such that the following conditions are fulfilled. (a) $|h^\mu| < k$ for all $\mu = 1, \ldots, p$, where k is the tolerance bound. We assume that the vectors ξ_j^μ are randomly distributed, with $\langle\langle \xi_j^\mu \rangle\rangle = 0$, and $\langle\langle \xi_i^\mu \xi_j^\nu \rangle\rangle = \delta_{ij}\delta_{\mu\nu}$. Hence, they introduce symmetric constraints to the solution space. (b) The normalization condition $\sum_j J_j^2 = N$. (c) $|\sum_j a_j J_j/\sqrt{N}| = 1$; however, this condition is easily satisfied: if there exists a solution satisfying (a) and (b) but yields $|\sum_j a_j J_j/\sqrt{N}|$ different from 1, it is possible to make an adjustment of each component J_j proportional to $\operatorname{sgn} a_j/\sqrt{N}$. Since the noise components ξ_j^μ are uncorrelated with a_j, the local fields make a corresponding adjustment of the order $1/\sqrt{N}$, which vanishes in the large-N limit. The space of the vectors J_j satisfying the constraints (a) and (b) is referred to as the *version space*.

NR is a prototype for a wider class of important problems such as load balancing in computer networks, traffic regulation and production management. A server, such as a computer server, a traffic junction or a factory, receives input quantities λ_j from source j (=1, …, N), which may be computer jobs, traffic flow or raw materials respectively. Input λ_j is distributed by the server to outputs μ (=1, …, p) with given probabilities $p(\mu|j)$, and the load received by output μ is $\sum_j \lambda_j p(\mu|j)$. The optimization task is to determine the inputs λ_j, so that the load forwarded to each output only deviates from the average within a tolerance. Letting $\lambda_j = \langle\lambda_j\rangle + J_j$ and $p(\mu|j) = p^{-1} + \xi_j^\mu$, and normalizing J_j, we obtain constraints (a) and (b).

Although very similar to the problem of pattern storage in the perceptron with continuous couplings where the constraints (a) are $h^\mu > k$ [9], the NR model has an extra inversion symmetry: the version space is invariant under $\vec{J} \to -\vec{J}$. The NR model is also a simplified version of the perceptron with multi-state output [10], in which the values of local fields for each pattern are bounded in one of the few possible intervals, and here we only have a single symmetric interval $[-k, k]$. Although they share the common feature that the version space is not connected or not convex, the extra symmetry will lead to very different phase behaviour from other perceptron models with, e.g., errors [11], discrete [12] or pruned couplings [13], or non-monotonic transfer functions [14].

When the number of noise sources increases, the version space is reduced and undergoes a sequence of phase transitions, causing it to disappear eventually. These transitions are observed by monitoring the evolution of the overlap order parameter q, which is the typical overlap between two vectors in the version space. For few noise sources, the version space is extended and $q = 0$. When the number of noise sources p increases, the number of constraints increases and the version space shrinks.

One possible scenario is that each constraint reduces the volume of the version space, and there is a continuous transition to a phase of non-zero value of q. Alternatively, each constraint introduces a volume reduction resembling a percolation process, in which the version space remains extended until a sufficient number of constraints have been introduced, and the version space is suddenly reduced to a localized cluster. This may result in a discontinuous transition from zero to non-zero q. We expect that the transition takes place when p is of the order N, and we define $\alpha \equiv p/N$ as the noise population. When α increases further, q reaches its maximum value of 1 at $\alpha = \alpha_c$, which is called the *critical population*. The purpose of this paper is to study the nature and conditions of occurrence of these transitions.

We calculate the entropy (i.e. the self-averaging logarithm of the volume of the version space). Averaging over the noise sources using the replica method [9], $\mathcal{S} = \lim_{n\to 0}(\langle\langle \mathcal{V}^n \rangle\rangle - 1)/n$, where

$$\langle\langle \mathcal{V}^n \rangle\rangle = \left\langle\!\!\left\langle \prod_{a=1}^{n} \int \prod_{j=1}^{N} \mathrm{d}J_j^a \delta\left(\sum_{j=1}^{N} J_j^{a2} - N \right) \prod_{\mu=1}^{p} \theta(k^2 - h_a^{\mu 2}) \right\rangle\!\!\right\rangle \tag{1}$$

with $h_a^\mu \equiv \sum_j J_j^a \xi_j^\mu / \sqrt{N}$. The result is $\langle\langle \mathcal{V}^n \rangle\rangle = \int \prod_{a<b=1}^{n} \mathrm{d}q_{ab} \exp(Ng)$. The overlaps between the coupling vectors of distinct replicas a and b: $q_{ab} \equiv \sum_{j=1}^{N} J_j^a J_j^b / N$, are determined from the stationarity conditions of g.

Due to the inversion symmetry of the constraints, it always has the all-zero solution ($q_{ab} = 0, \forall a < b$), but it becomes locally unstable at a noise population

$$\alpha_{\mathrm{AT}}(k) = \frac{\pi}{2} \frac{\mathrm{erf}(k/\sqrt{2})^2}{k^2 \exp(-k^2)}. \tag{2}$$

For $\alpha > \alpha_{\mathrm{AT}}$, the simplest solution assumes $q_{ab} = q > 0$. This replica symmetric (RS) solution, however, is not stable against replica symmetry breaking (RSB) fluctuations for any $q > 0$. Hence, (2) is an Almeida–Thouless (AT) line [1], and RSB solutions in the Parisi scheme [1] have to be considered.

The transition of q from zero to non-zero is absent in the problem of pattern storage in the perceptron, where q increases smoothly from zero when the storage level α increases [9]. Rather, the situation is reminiscent of the spin glass transition in the Sherrington–Kirkpatrick (SK) model, which does possess an inversion symmetry [1]. The phase diagram is now discussed in the following three schemes (technical details will be published elsewhere).

(1) The RS solution (RS, superscript $^{(0)}$) is given by $q = q_{\mathrm{EA}}$, where q_{EA} is the Edwards–Anderson order parameter [1]. Close to the AT line, $q \sim t$, where $t \equiv (\alpha - \alpha_{\mathrm{AT}})/\alpha_{\mathrm{AT}} \ll 1$. The critical population ($q \to 1$) is

$$\alpha_c^{(0)}(k) = \left((1 + k^2) \left(1 - \mathrm{erf}\left(\frac{k}{\sqrt{2}} \right) \right) - \sqrt{\frac{2}{\pi}} k \exp\left(\frac{-k^2}{2} \right) \right)^{-1}. \tag{3}$$

Two features are noted in the phase diagram in figure 1:

(a) The critical population line crosses the AT line. When $k = 0$, the version space is equivalent to the solution of p homogeneous N-dimensional linear equations. Hence it vanishes at $p = N$, or $\alpha_c^{(0)} = 1$. For small k, $\alpha_{\mathrm{AT}}(k)$ increases from 1 quadratically, and is therefore less than $\alpha_c^{(0)}(k)$, which increases from 1 linearly. Thus q increases smoothly from 0 at α_{AT} to 1 at $\alpha_c^{(0)}$. However, in the large-k limit, $\alpha_{\mathrm{AT}}(k)$ grows exponentially with k^2, whereas $\alpha_c^{(0)}(k)$ grows exponentially with $k^2/2$ only, and the reverse is true.

(b) The first-order transition line takes over the AT line. The paradox in (a) is resolved by noting that for a given k, multiple RS solutions of q may coexist for a given α. Indeed, $\alpha^{(0)}(k, q)$ is monotonic in q for $k \leqslant k_c^{(0)} = 2.89$, and not so otherwise. Among the coexisting stable RS solutions for $k > k_c^{(0)}$, the one with the lowest entropy is relevant. Hence, there is a first-order transition in q at a value $\alpha_1^{(0)}(k)$, determined by the vanishing of the entropy difference between the coexisting stable solutions. The jump in q across the transition widens from 0 at $k = k_c^{(0)}$, and when k increases above $k_0^{(0)} = 3.11$, q jumps directly from zero to non-zero at the transition point. The line of $\alpha_1^{(0)}(k)$ starts from $k = k_c^{(0)}$; it crosses the line of $\alpha_{\mathrm{AT}}(k)$ at $k_0^{(0)}$, replacing it to become the physically observable phase transition line. In the limit of large k, the first-order transition line is given by $\alpha_1^{(0)}(k) = 0.94\alpha_c^{(0)}(k)$, where q jumps from 0 to a high value of $1 - 0.27k^{-2}$.

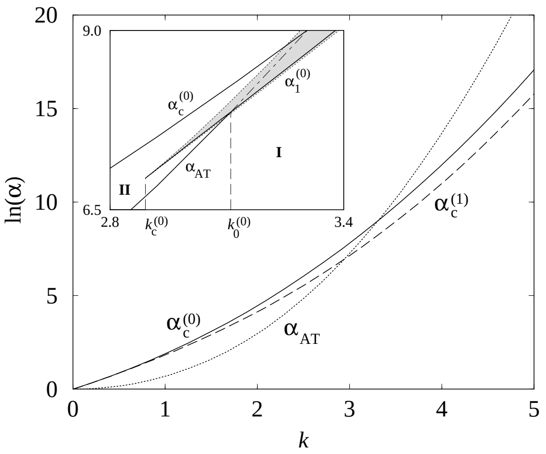

Figure 1. Comparison of $\alpha_c^{(0)}$, $\alpha_c^{(1)}$ and $\alpha_{\rm AT}$. Inset: the picture within RS near $k_c^{(0)}$ and $k_0^{(0)}$, with two areas: (I) $q = 0$ and (II) $q > 0$. Shaded area: multiple solutions of q.

These observations illustrate the nature of the phase transitions. For low tolerance k, each constraint results in a significant reduction in the version space, and there is a continuous transition to a phase of non-zero value of q. For high tolerance k, each constraint introduces a volume reduction which is less significant, resembling a percolation process, in which the transition of q from zero to non-zero is discontinuous. This picture is refined in the next approximation.

(2) In the first step RSB approximation (RSB$_1$, superscript $^{(1)}$) the n replicas are organized into clusters of m replicas. $q_{ab} = q_1$ for replicas in the same cluster, and $q_{ab} = q_0$ otherwise, and in the limit $n \to 0$, $0 < m < 1$ for analytic continuation.

Two (q_1, q_0, m) solutions exist just above the AT line: one with $q_0 = 0$ and one with $q_0 > 0$ (see figure 3). Only the latter is stable with respect to fluctuations of q_0.

The features in the phase diagram in figure 2 are as follows.

(a) First-order transition: For $k > k_c^{(1)} = 2.31$ multiple solutions exist, and there is a first order transition of q at $\alpha_1^{(1)}(k)$. This line starts from $k = k_c^{(1)}$, and crosses the line of $\alpha_{\rm AT}(k)$ at $k_0^{(1)} = 2.61$, to become the physically observable phase transition line. For large k, $q_1 = 1 - 1.23/(k \ln k)^2$ at the transition, and $\alpha_1^{(1)}(k) = (1 - 0.33/\ln k)\alpha_c^{(1)}(k)$. Hence, the system first undergoes a percolation transition, in which the occurence of localized clusters is indicated by the discontinuous transition from zero to non-zero q_1.

(b) Reduced critical population *compared with* $\alpha_c^{(0)}(k)$. In the large-k limit, $\alpha_c^{(1)}(k)/\alpha_c^{(0)}(k) \simeq 2\ln k/k^2$.

(c) Transition of q_0. The regime of non-zero q_1 and q_0 spans the region of low k below the critical population. At intermediate values of k, however, there is a line $\alpha_m^{(1)}$ where q_0 becomes zero. The line starts from $k_{c,m}^{(1)} = 2.37$ on the line of critical population, and ends at $k = k_0^{(1)} = 2.61$ on the line of first-order transition. Beyond this line in the localized phase, only the solution with $q_0 = 0$ exists, reflecting the fact that the distribution of clusters remains extended and isotropic.

The formation of localized clusters at the percolation transition is illustrated by the evolution of the overlap distribution. For $k > k_0^{(1)}$ at the first-order transition line, the parameter m decreases smoothly, with increasing α, from 1 across the phase transition line.

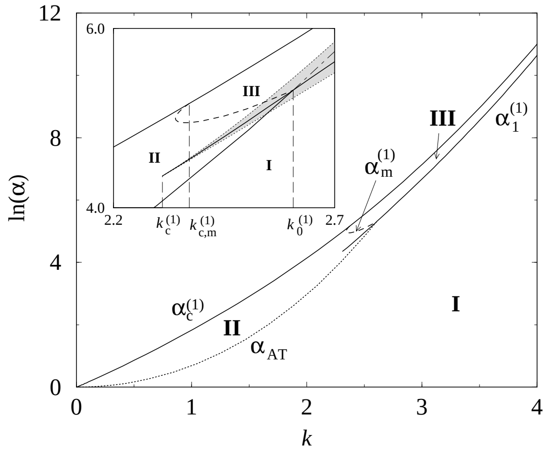

Figure 2. The picture within RSB_1, with the three areas: (I) $q_0 = q_1 = 0$, (II) $0 < q_0 < q_1 < 1$ and (III) $0 < q_1 < 1$, $q_0 = 0$. Inset: the region near $k_c^{(1)}$ and $k_0^{(1)}$. Shaded area: multiple solutions of q_1 and q_0.

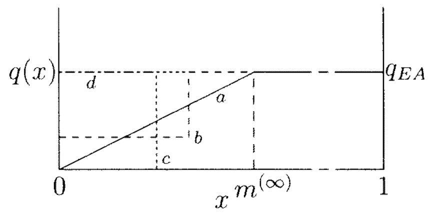

Figure 3. The RSB_∞ solution near the AT line (*a*). For comparison, the RSB_1 solution ((*b*): non-zero q_0, (*c*): zero q_0) and the RS solution (*d*) are also plotted.

This is analogous to the RS–RSB_1 transition in the random energy model [1, 15] and the high-temperature perceptron [16]. Since the overlap distribution is $P(q) = m\delta(q - q_0) + (1 - m)\delta(q - q_1)$, the transition is continuous in the overlap distribution, although discontinuous in terms of q_1. At the transition, the statistical weight of the $q = q_0 = 0$ component decreases smoothly, while that of the $q = q_1$ component increases from zero. Two points in the version space have a high probability to share an overlap q_0, meaning that they belong to different clusters. Hence, the version space consists of many small clusters, which are isotropically distributed, since $q_0 = 0$. When α increases, m decreases towards 0, and the version space is mainly reduced by weeding out clusters.

In contrast, for the transition at low k, the overlap has the same form as the SK model, i.e. a high probability being q_1, implying that it consists of few large clusters.

(3) In the infinite step RSB solution (RSB_∞, superscript $^{(\infty)}$) the n replicas are organized into hierarchies of clusters m_i. In the Parisi scheme [1], the overlap q_{ab} is represented by the Parisi function $q(x)$, where x is the cumulative frequency of q, or $P(q) = \mathrm{d}x(q)/\mathrm{d}q$. We have only obtained solutions for α just above α_{AT}. As shown in figure 3, $q(x)$ grows linearly from $x = 0$ to $x = m^{(\infty)}$ and remains constant at $q_{\mathrm{EA}} \sim t$ until $x = 1$. It is very similar to the Parisi function of the SK model without external field near the critical temperature [1].

For the RSB_∞ solution far away from the AT line, we only give some qualitative features of the phase behaviour. Below the AT line we have the extended phase characterized by $q(x) = 0$. For small k, $\alpha_{\mathrm{AT}}(k)$ is less than $\alpha_c^{(\infty)}(k)$. Hence for α increasing above α_{AT}, there is a continuous transition of $q(x)$, and $q_{\mathrm{EA}} \to 1$ for $\alpha \to \alpha_c^{(\infty)}(k)$.

For large k, $\alpha_{\mathrm{AT}}(k)$ is greater than $\alpha_c^{(\infty)}(k)$, since $\alpha_c^{(\infty)}(k) \leqslant \alpha_c^{(1)}(k) < \alpha_c^{(0)}(k)$. Hence, the $\alpha_c^{(\infty)}(k)$ line must intersect the AT line, and there must exist a critical $k_c^{(\infty)}$ above which there is a first-order transition from a solution with low (or zero) q_{EA} to one with high q_{EA} (denoted by $\alpha_1^{(\infty)}(k)$). This line intersects the $\alpha_{\mathrm{AT}}(k)$ line at a value $k_0^{(\infty)}$, replacing it to be the physically observable phase transition line.

We expect the picture of version space percolation for large k also to be valid in the RSB_∞ ansatz. When α increases above $\alpha_1^{(\infty)}(k)$, $q(x)$ will deviate appreciably from the zero function only in a narrow range of x near 1. Hence, the overlap q is zero with a probability almost equal to 1, and non-zero with a probability much less than 1. The corresponding picture is that the version space consists of hierarchies of localized clusters which are scattered in all directions of an N-dimensional hypersphere. Again, the transition is continuous in terms of the overlap distribution, though not so in terms of $q(x)$.

The phase transitions observed here have strong implications to the NR problem. In the extended phase for low values of α, an adaptive algorithm can find a solution easily in any direction of the N-dimensional parameter space. In the localized phase, the solution can only be found in certain directions. If the tolerance k is sufficiently high, there is a jump in the overlap distribution, which means that the search direction is suddenly restricted on increasing α. For very high tolerance k, the picture of percolation transition applies, so that even though localized clusters of solutions are present in all directions, it is difficult to move continuously from one cluster to another without violating the constraints.

The analysis of NR can be extended to perceptrons with *multi-state* and *analogue* outputs, when there are sufficiently many intermediate outputs. The version space formed by the intermediate states has the same geometry as the model studied here. To picture this, note that for $\alpha > 1$, the part of space delimited by the constraints (a), forms a random closed polytope (box), constraint (b) is the surface of a hypersphere, and the version space is their intersection. Hence for sufficiently large p, it consists of disconnected clusters, resulting in the occurence of RSB [10, 17]. Percolation-like transitions are expected in the case of transfer functions of low gain.

In conclusion, we have introduced a model for optimization problems which is simple, calculable within the full Parisi RSB scheme and shows rich non-trivial behaviour. It can serve as a prototype for a wide class of problems in disordered systems such as NR in signal processing, traffic flow, load balancing, information storage in multi-state and analogue neural networks. Moreover, it explains the possible phase transitions in the solution space for these problems, where both volume reduction and percolation-like scenarios are possible. This is an important consideration prior to the construction of any optimal optimization algorithms.

We thank R Kühn for informative discussions, M Bouten, T Coolen, H Nishimori and P Ruján for critical comments. This work is partially supported by the Research Grant Council of Hong Kong (HKUST6130/97P), by the Research Fund of the K U Leuven (grant OT/94/9) and by the EPSRC (GR/M11554).

References

[1] Mézard M, Parisi G and Virasoro M 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[2] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Redwood City, CA: Addison-Wesley)
[3] Sourlas N 1989 *Nature* **339** 693
[4] Monasson R and Zecchina R 1996 *Phys. Rev. Lett.* **76** 3881
Monasson R and Zecchina R 1997 *Phys. Rev.* E **56** 1357
[5] Mertens S 1998 *Phys. Rev. Lett.* **81** 4281
[6] Monasson R, Zecchina R, Kirkpatrick S, Selman B and Troyanski L 1999 *Nature* **400** 133
[7] Haykin S 1991 *Adaptive Filter Theory* 2nd edn (Englewood Cliffs, NJ: Prentice-Hall)
[8] Jutten C and Herault J 1991 *Signal Process.* **24** 1
Basak J and Amari S 1999 *Neural Comput.* **11** 1011
[9] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[10] Bollé D and van Mourik J 1994 *J. Phys. A: Math. Gen.* **27** 1151
[11] Erichsen Jr R and Theumann W K 1993 *J. Phys. A: Math. Gen.* **26** L61
[12] Krauth W and Mézard M 1987 *J. Physique* **50** 3057

[13] Garcés R, Kuhlman P and Eissfeller H 1992 *J. Phys. A: Math. Gen.* **25** L1335
[14] Bollé D and Erichsen R 1996 *J. Phys. A: Math. Gen.* **29** 2299
[15] Fischer K H and Hertz J 1991 *Spin Glasses* (Cambridge: Cambridge University Press)
[16] Györgyi G and Reimann P 1997 *Phys. Rev. Lett.* **79** 2746
[17] Bollé D, Kühn R and van Mourik J 1993 *J. Phys. A: Math. Gen.* **26** 3149