Optimization of output load in a junction model

2001 J. Phys. A: Math. Gen. 34 L635

(http://iopscience.iop.org/0305-4470/34/44/106)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **34** (2001) L635–L642 PII: S0305-4470(01)15353-9

LETTER TO THE EDITOR

Optimization of output load in a junction model

K Y Michael Wong[1] **and Fu-li Li**[2]

[1] Department of Physics, Hong Kong University of Science and Technology, Hong Kong, People's Republic of China
[2] Department of Applied Physics, Xian Jiaotong University, Xian 710049, People's Republic of China

Received 10 July 2000, in final form 18 September 2001
Published 26 October 2001
Online at stacks.iop.org/JPhysA/34/L635

Abstract
An optimization problem of minimizing the output load of a junction model with multiple input and output channels is solved by the replica method. Simulation results show that for the case of extensive connectivity, the replica solution obtained by assuming a vanishing solution space is satisfactory. However, for the case of intensive connectivity, the degeneracy of the solution space causes the simulation results to deviate from the replica solution of vanishing volume.

PACS numbers: 64.60.Cn, 89.70.+c, 02.50.-r, 05.20.-y

An interesting topic in statistical mechanics is its applications to the study of optimization involving an extensive number of variables. To mention a few examples, we have the graph matching problem [1], the graph bipartitioning problem [2], the storage capacity of perceptrons [3, 4], the K-satisfiability problem [5–7], the number partitioning problem [8] and the noise reduction model [9]. In many cases, phase transitions were observed in these systems, and the entropy at the critical point can be computed. The replica method [10], which was developed from the study of spin glasses and extended to deal with systems involving a large number of interacting variables [11, 12] has proved to provide an effective approach to these problems.

In this paper we will apply the replica method to study the load optimization in a junction model with randomly connected multiple inputs and outputs, and compute the minimum upper bound of the load. This model can be applied to load balancing in computer networks, traffic control and production line management.

While it is very interesting to consider the global optimization of the load network-wide, its applications in networks of increasing sizes (such as the Internet) are increasingly difficult. Hence there is a considerable shift of interest to distributed and local optimization recently, and the present model of a single junction is thus important and relevant. Furthermore, if one were to understand the network-wide behaviour, the properties of a single junction needs to be studied first.

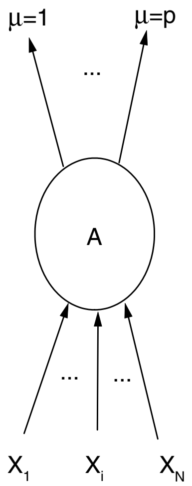

Figure 1. A junction model with multiple input and output channels.

As shown in the junction model in figure 1, the junction A has N input and p output channels. Both $p \gg 1$ and $N \gg 1$, and the number ratio of the output and input channels is $\alpha = pN^{-1}$. Suppose the load on the jth input x_j is normalized by $\sum_j x_j = N$. The junction is fed by the load from each of the inputs, and allocates them to all the outputs. Let $p(\mu|j)$ represent the fraction of load allocated from the jth input to the μth output, satisfying the normalization condition $\sum_\mu p(\mu|j) = 1$. Then the total load on the μth output is $\sum_j p(\mu|j)x_j$.

In many applications, load optimization often requires the junction to adjust the input load x_j so that the load on the outputs is as uniform as possible. The optimization of traffic through a tandem in telecommunications networks belongs to this kind of problem [13]. This can be formulated as a linear programming problem. Specifically, one can minimize the upper bound B of the load on the outputs, where B satisfies the condition $\sum_j p(\mu|j)x_j \leqslant B$.

When the problem under investigation involves a large number of inputs and outputs, the method of statistical physics becomes very useful. For the present case, let us consider a specific distribution of the probability $p(\mu|j)$. The average value of $p(\mu|j)$ is p^{-1}. Suppose that for a given output channel μ, there corresponds an average of $C \gg 1$ input channels whose distributed load experience fluctuations about the background average of p^{-1}. For the rest of the input channels feeding μ, the distributed load is maintained at the uniform level of p^{-1}. To enable comparison between situations of different connectivities C, we assume that the fluctuating load has fractional fluctuations of the order $C^{-1/2}$. Specifically, suppose $p(\mu|j) = (1 + C^{-1/2}\xi_j^\mu)p^{-1}$, where ξ_j^μ takes a zero value with probability $1 - CN^{-1}$. With probability CN^{-1}, it takes a nonzero random value with mean 0 and variance 1. Since the typical value of ξ_j^μ is of the order 1, the probability that p_j^μ becomes negative is negligible for $C \gg 1$. Therefore, without optimizing, the mean load on the μth output is $\sum_j p(\mu|j)x_j = \alpha^{-1}$. In general, the upper bound of the load may be larger than α^{-1} because the random allocation in the junction induces fluctuations on the outputs. The purpose of the optimization is to minimize the upper bound.

The randomness of the nonzero allocation probability affects the optimization result. The load on the μth output is $\alpha^{-1} + (p\sqrt{C})^{-1} \sum_j' x_j \xi_j^\mu$, where $\sum'$ represents the summation over the inputs with nonzero values of ξ_j^μ. The upper bound of the load can be written as $B = \alpha^{-1} + bp^{-1}$. Hence the original problem can be described by the following linear optimization

$$\text{minimize } b\text{, subject to}$$

$$\frac{1}{\sqrt{C}}\sum_{j=1}^{N}\xi_j^\mu x_j \leqslant b \qquad \mu = 1, 2, \ldots, p \tag{1}$$

$$\sum_{j=1}^{N} x_j = N \tag{2}$$

$$x_j \geqslant 0. \tag{3}$$

This formulation is similar to the problem of storage capacity in neural networks [3,4] and the noise reduction model [9].

Note that for $p \sim N \gg 1$, b can be either positive or negative, and the probability that B becomes negative is negligible since, as we shall see, $b \sim N^0$. Hence the objective of optimization is to minimize the upper bound of the fluctuating component of the load, while the average load itself is maintained at α^{-1}. While the fluctuating components themselves are relatively small, their optimization is still a critical issue when, for instance, the values of the input load x_j themselves represent the averages of time-varying traffic. This is the case in telecommunications networks when the output channels are nearly saturated by the output traffic; time-varying load may occasionally block the channels completely, leading to a degradation in service. Minimizing the fluctuations will therefore minimize the blocking probabilities [13].

When N goes to infinity while C remains finite, the problem belongs to a class of infinite-ranged problems in which the local connection topology can be mapped onto tree structures [14]. In such problems the thermodynamic variables per node (such as the free energy per node or the entropy per node) are functions of the connectivity C. In other words, the finite value of C remains relevant, even when C does not scale with N. Such tree-like approximations found applications in many problems of practical interest, including graph bipartitioning [14], diluted neural networks [15], computational logic [5–7] and error-correcting codes [16].

In the present context, each output μ is fed by an average of C inputs, each of which in turn feeds an average of αC outputs. This maps onto a graph with nodes representing inputs and outputs, and lines between nodes with connecting fluctuating traffic. The probability of finding a loop of finite length m scales as $C^m N^{-m}$. Since this is negligible in the limit that $1 \ll C \ll N$, the local topology is tree like.

Here, we will use the replica method to find the optimal value $b_{\min}$ for typical cases. The result will be compared with that obtained by numerical simulation experiments using the standard linear programming approach.

Consider the volume V in the x-space defined by conditions (1)–(3). For a given realization of $\{\xi_j^\mu\}$, V can be written as

$$V = \int_0^\infty \prod_{j=1}^{N} \mathrm{d}x_j\, \delta\bigg(\sum_{j=1}^{N} x_j - N\bigg) \prod_{\mu=1}^{P} \Theta\bigg(b - \frac{1}{\sqrt{C}}\sum_{j=1}^{N}\xi_j^\mu x_j\bigg) \tag{4}$$

where Θ is the step function. We consider the extensive quantity $\ln V$, i.e. the entropy of the solution space, and calculate its ensemble average using the replica method [10–12],

$$\langle \ln V\rangle = \lim_{n\to 0}\frac{\langle V^n\rangle - 1}{n} \tag{5}$$

where $\langle\rangle$ represents averaging over the sampling of the coefficients ξ_j^μ. From (4), we have

$$\langle V^n\rangle = \int_0^\infty \prod_{\alpha=1}^{n}\prod_{j=1}^{N} \mathrm{d}x_j^\alpha\, \delta\left(\sum_{j=1}^{N} x_j^\alpha - N\right)\left\langle \prod_{\alpha=1}^{n}\prod_{\mu=1}^{p}\Theta\left(b-\frac{1}{\sqrt{C}}\sum_{j=1}^{N}\xi_j^\mu x_j^\alpha\right)\right\rangle \tag{6}$$

where index α denotes the αth replica of the original system. In the limit of $N\to\infty$, the result is

$$\langle V^n\rangle = \exp[NG(E,q,\hat{q})] \tag{7}$$

where

$$\begin{aligned}G(E,q,\hat{q}) = {} & \sum_\alpha E_\alpha + \sum_{r_1\cdots r_n}\frac{1}{r_1!\cdots r_n!}\hat{q}_{r_1\cdots r_n}q_{r_1\cdots r_n} + \ln\prod_\alpha\int_0^\infty \mathrm{d}x_\alpha \\ & \times \exp\left[-\sum_\alpha E_\alpha x_\alpha - \sum_{r_1\cdots r_n}\frac{1}{r_1!\cdots r_n!}\hat{q}_{r_1\cdots r_n}(x_1)^{r_1}\cdots(x_n)^{r_n}\right] \\ & +\alpha\ln\prod_\alpha\int_{-\infty}^{b}\mathrm{d}\lambda_\alpha\int\frac{\mathrm{d}\hat{\lambda}}{2\pi}\exp\left[\sum_\alpha \mathrm{i}\hat{\lambda}_\alpha\lambda_\alpha\right. \\ & \left. +C\int\rho(\xi)\sum_{r_1\cdots r_n}\frac{1}{r_1!\cdots r_n!}\left(-\frac{\mathrm{i}\xi}{\sqrt{C}}\right)^{r_1+\cdots+r_n} q_{r_1\cdots r_n} - C\right].\end{aligned} \tag{8}$$

Here q is the order parameter

$$q_{r_1r_2\cdots r_n} = \frac{1}{N}\sum_{j=1}^{N}(x_j^1)^{r_1}(x_j^2)^{r_2}\cdots(x_j^n)^{r_n} \tag{9}$$

E and $\hat{q}$ are the Lagrange multipliers of constraints (2) and (9) respectively. They are determined from the saddle point equations of (7). If the distribution of random variables ξ_j^μ is an even function, then $\langle\xi^{\sum_\alpha r_\alpha}\rangle = 0$ for $\sum_\alpha r_\alpha =$ odd, $q_{r_1r_2\cdots r_n} = \hat{q}_{r_1r_2\cdots r_n} = 0$.

In general, it is difficult to find solutions to the saddle point equations of (7). However, a particularly simple solution exists when $C \gg 1$, for which we need only to keep components of variables q and $\hat{q}$ with $\sum_\alpha r_\alpha = 2$. All the other components of $\hat{q}$ are of orders C^{-1} or higher and can be neglected, and (7) becomes independent of their corresponding order parameters. Introducing the replica symmetric ansatz [10–12], there are now five distinct parameters:

(a) q_2 and its corresponding Lagrange multiplier $\hat{q}_2$, where q_2 is the average value of the mean x_j^2 in the solution space, i.e.

$$q_2 = \frac{1}{N}\sum_j\langle x_j^2\rangle. \tag{10}$$

(b) q_{11} and its corresponding Lagrange multiplier $\hat{q}_{11}$, where q_{11} is the average value of the squared mean x_j in the solution space, i.e.

$$q_{11} = \frac{1}{N}\sum_j\langle x_j\rangle^2. \tag{11}$$

(c) E is the Lagrange multiplier of constraint (2).

In the present optimization problem, we are interested in the condition of taking the minimal value for b. When $C \gg 1$, each input is connected to the many outputs and the solution space is not degenerate. Therefore, the volume of the solution space shrinks to zero as

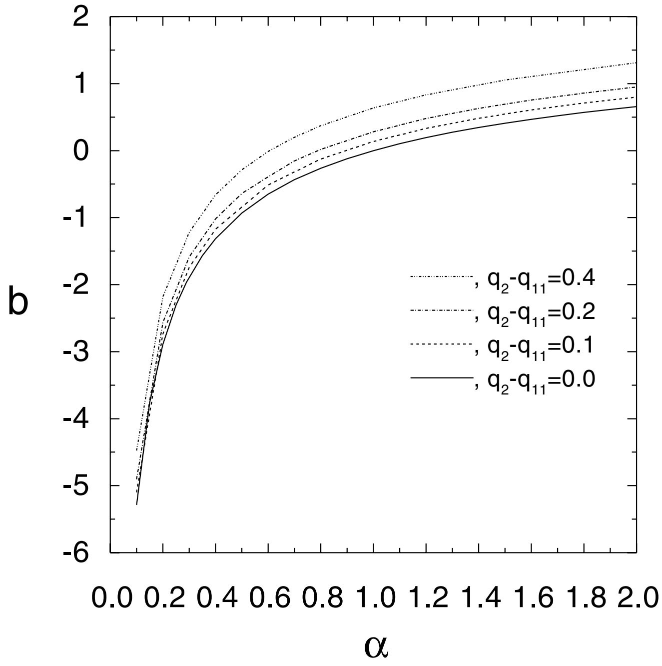

Figure 2. b versus α for given values of $q_2 - q_{11}$. The solid curve represents the result obtained from equations (12)–(14).

b approaches the minimal value. In this case, one should have $q_2 - q_{11} \to 0$ and the conjugate $\hat{q}_2 - \hat{q}_{11} \to \infty$. In the limit $n \to 0$, the saddle point equations of (7) become

$$\alpha H\left(\frac{b_{\min}}{\sqrt{q_{11}}}\right) = H\left(\frac{E}{\sqrt{\hat{q}_{11}}}\right) \tag{12}$$

$$q_{11}\left[\int_{E/\sqrt{\hat{q}_{11}}}^{\infty} Dt\left(t - \frac{E}{\sqrt{\hat{q}_{11}}}\right)\right]^2 = \int_{E/\sqrt{\hat{q}_{11}}}^{\infty} \mathrm{D}t\left(t - \frac{E}{\sqrt{\hat{q}_{11}}}\right)^2 \tag{13}$$

$$\int_{b_{\min}/\sqrt{q_{11}}}^{\infty} \mathrm{D}t\left(t - \frac{b_{\min}}{\sqrt{q_{11}}}\right)^2 \int_{E/\sqrt{\hat{q}_{11}}}^{\infty} \mathrm{D}t\left(t - \frac{E}{\sqrt{\hat{q}_{11}}}\right)^2 = H\left(\frac{E}{\sqrt{\hat{q}_{11}}}\right) H\left(\frac{b_{\min}}{\sqrt{q_{11}}}\right) \tag{14}$$

where

$$H(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \mathrm{e}^{-y^2/2}\,\mathrm{d}y. \tag{15}$$

As far as the above result is concerned, the random variable ξ_j^μ may be either the standard Gaussian or the discrete Ising variable. Only when $C \sim O(1)$ does the result depend on the specific distribution of ξ_j^μ.

Figure 2 shows the values of b versus α for given values of $q_2 - q_{11}$. We observe that $b_{\min}$ increases from negative to positive with α, passing through the point $(\alpha, b_{\min}) = (1, 0)$. When α approaches zero, $b_{\min}$ approaches $-\sqrt{2|\ln\alpha|/\alpha}$. When $\alpha \gg 1$, $b_{\min}$ approaches $\sqrt{2\ln\alpha}$. To illustrate the effects of optimization, we may compare this result with the upper bound before optimization. In the unoptimized case, the load of the outputs in (1) are Gaussian

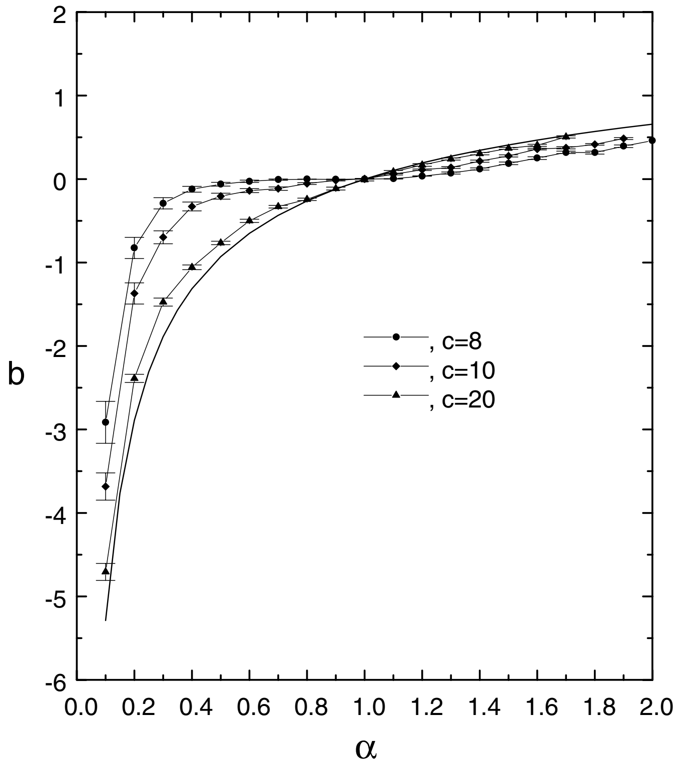

Figure 3. Results of numerical simulations for the relation of α and b, obtained for $N = 200$ and different values of C. The solid curve represents the result obtained from (12)–(14).

variables with mean 0 and variance 1. If U is the upper bound of p Gaussian variables, then its probability is given by

$$P(U) = p\frac{\exp(-U^2/2)}{\sqrt{2\pi}}\,(H(-U))^{p-1}\,. \qquad (16)$$

For large p, the peak of $P(U)$ is located at $U \sim \sqrt{2\ln p}$, which is much larger than the optimized bounds.

To check the validity of (12)–(14), simulations are run using the simplex method [17]. Results for $N = 200$ are shown in figure 3. We can see that when C is large, the simulation results agree with the predictions of (12)–(14). However, when C decreases, significant deviations from the theory is observed. It is interesting to note that the optimal value of $b_{\min}$ is essentially 0 over a wide parameter range, whereas the theory predicts that $b_{\min} = 0$ only at $\alpha = 1$.

This phenomenon can be understood as a consequence of the degeneracy of the solution space. Consider the matrix elements ξ_j^μ. For a given row μ, there are on average C nonzero elements. Similarly, for a given column j, there are on average αC nonzero elements. Since the choice of nonzero elements is random, the number of nonzero elements per row is Poisson distributed. Hence with probability e^{-C}, all elements in row μ are zero. Similarly, with probability $\mathrm{e}^{-\alpha C}$, all elements in column j are zero. When C becomes of the order 1, these probabilities cannot be neglected. Hence all-zero rows or columns are present.

When row μ is all-zero, its corresponding constraint (1) becomes $0 \leqslant b$. Hence $b_{\min}$ is bounded below by 0. In this case, the solution space can be seen to be degenerate by considering the $N + 1$ dimension space formed by $x_1, \ldots, x_N, b$. If the superspace formed by all constraints except the μth one has a negative upper bound, then the corresponding solution space is a convex polygon with a vertex at a negative value of b and extending to positive infinity in b. Now if the μth constraint is included, this vertex and its vicinity is 'chopped off' by the plane $b = 0$. $b_{\min}$ is now 0, and the solution space is degenerate. This explains the simulation result of a near-zero value of $b_{\min}$ for the range of α less than 1.

When column j is all-zero, x_j does not appear in any of the p constraints (2). In this case, the solution space contains the point $x_i = N\delta_{ij}$, corresponding to an upper bound of $b = 0$. Hence $b_{\min}$ cannot be greater than 0. When there are more than one all-zero column $j_1, j_2, \ldots$, the solution space corresponding to $b = 0$ is given by a region in the plane $x_{j_1} + x_{j_2} + \cdots = N$, forming a degenerate space. This explains the simulation results of a near-zero value of $b_{\min}$ for the range of α greater than 1.

The optimization problem discussed here is a problem in linear programming. Its solution space is convex. Hence the optimal value of the upper bound is globally optimal. No local minima are present, and the replica-symmetric ansatz is valid. On the other hand, if one allows the violation of some constraints, a replica symmetry-breaking solution may be necessary.

To summarize, we have found that the replica theory embodied in equations (12)–(14) breaks down when the connectivity C becomes intensive. The main reason is that the solution space is degenerate, and its volume remains finite. Hence the assumption that $q_2 - q_{11}$ vanishes is not valid. For a given value of α, the entropy of the solution space decreases to a finite value when b approaches $b_{\min}$, and suddenly vanishes below $b_{\min}$. This is reminiscent of the first-order transition in the K-satisfiability problem [5]. In contrast, the entropy diverges to $-\infty$ when $C \gg 1$, which is more analogous to the second-order transition of the storage capacity in neural networks [4].

In fact, large degrees of freedom of the solution space are common features of many optimization problems with intensive connectivity, as illustrated by the nonvanishing entropy in the graph bipartitioning problem [18] and the K-satisfiability problem [5]. These cases are very different from the cases of extensive connectivity such as the storage capacity in neural networks [3,4], and should be approached using separate methodologies and techniques. Since the junction model is a prototype of many optimization problems which undergo phase transitions, we anticipate that statistical mechanics will be a useful tool in these issues.

This work was supported by the grant HKUST6130/97P of the Research Grant Council of Hong Kong and the National Natural Science Foundation of China.

References

[1] Mézard M and Parisi G 1985 *J. Physique Lett.* **46** L771
[2] Fu Y and Anderson P W 1986 *J. Phys. A: Math. Gen.* **19** 1605
[3] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[4] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[5] Monasson R and Zecchina R 1996 *Phys. Rev. Lett.* **76** 3881
[6] Monasson R and Zecchina R 1997 *Phys. Rev.* E **56** 1357
[7] Monasson R, Zecchina R, Kirkpatrick S, Selman B and Troyanski L 1999 *Nature* **400** 133
[8] Mertens S 1998 *Phys. Rev. Lett.* **81** 4281
[9] van Mourik J, Wong K Y M and Bollé D 2000 *J. Phys. A: Math. Gen.* **33** L53
[10] Edwards S F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965
[11] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792
[12] Kirkpatrick S and Sherrington D 1978 *Phys. Rev.* A **17** 1384

[13] Lor W K F and Wong K Y M 1995 *Proc. Int. Workshop on Applications of Neural Networks to Telecommunications* ed J Alspector, R Goodman and T X Brown (Hillsdale: Erlbaum) p 137
[14] Wong K Y M and Sherrington D 1987 *J. Phys. A: Math. Gen.* **20** L793
[15] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
[16] Kabashima Y and Saad D 1999 *Europhys. Lett.* **45** 97
[17] Best M J and Ritter K 1985 *Linear Programming: Active Set Analysis and Computer Programs* (Englewood Cliffs, NJ: Prentice-Hall)
[18] Wong K Y M and Sherrington D 1988 *J. Phys. A: Math. Gen.* **21** L99