Replicas and optimization

M. Mézard (*) and G. Parisi (**)

Dipartimento di Fisica, Università di Roma I, Piazzale A. Moro 2, I-00185 Roma, Italy

Résumé. — Nous utilisons la méthode des répliques pour étudier le problème du matching (bipartite) avec des distances aléatoires indépendantes entre les points. Nous proposons une solution, symétrique dans les répliques, qui est en accord avec les valeurs numériques de la longueur minimale et la distribution des longueurs des liens occupés dans la configuration optimale.

Abstract. — We use the replica method to study the (bipartite) weighted matching problem with independent random distances between the points. We propose a replica symmetric solution which fits the numerical values of the minimal length and the distribution of lengths of the occupied links in the optimal configuration.

It has been noticed for some time that combinatorial optimization problems can be formulated as problems in statistical mechanics [1]. For a given instance of an optimization problem, one introduces an artificial temperature $T = 1/\beta$ and a Boltzmann weight on the various possible configurations C given by $e^{-E(C)}$ where $E(C)$ is the cost of the configuration, i.e. the quantity one wants to minimize. $E(C)$ can be called an energy, and in the language of statistical physics one looks for zero temperature properties of the system such as the ground state configuration and its energy.

Physically, the difficulties of optimization problems can be traced back to the absence of translational symmetry and the frustration (in the general sense that in the hard cases it is not possible to get the true ground state by a simple local optimization procedure). These kinds of complications are typically encountered in the statistical physics of disordered systems such as spin glasses, and the recent developments in this field, especially on the infinite range spin-glass [2] (which is an NP complete problem [3]) have risen the hope of a new approach to optimization problems.

Up to now, this approach has been mostly numerical. It has been shown that the Monte Carlo method, combined with a suitable annealing procedure, can be efficiently applied to some optimization problems [1, 4], and that the low lying metastable states (local minima of the cost function of low cost) possess some properties which are reminiscent of those of pure equilibrium states in spin glasses [5, 6].

(*) On leave from Laboratoire de Physique Théorique de l'Ecole Normale Supérieure, Paris.
(**) Università di Roma II, Tor Vergata, Roma and INFN, Sezione di Frascati.

In this paper we will show how the analytical methods developed in the mean field theory of spin glasses, and particularly the replica method, can be used to solve optimization problems. This being a kind of a pioneering paper, we have decided to present the method on a rather simple problem (a polynomial one) the weighted matching [7]. In this problem one is given $2N$ points $i = 1, ..., 2N$, with a matrix of distances l_{ij}, and one looks for a matching between the points (a set of N links between two points such that at each point one and only one link arrives) of minimal length.

An instance of this problem is a matrix l_{ij}. Hereafter we shall study the random-link case introduced in [5, 8] where l_{ij} are independent random variables identically distributed with a distribution $\rho(l)$. The configurations will be conveniently characterized by the set of the occupation numbers of the links, $n_{ij} = 0, 1$, which are subject to the constraints :

$$n_{ij} = n_{ji}; \quad \sum_{j=1}^{2N} n_{ij} = 1 \quad (1 \leqslant i \leqslant 2N).$$ (1)

The energy of a configuration is :

$$E(\{ n_{ij} \}) = \sum_{1 \leqslant i < j \leqslant 2N} n_{ij} \, l_{ij}$$ (2)

and the partition function is defined as :

$$Z = \sum_{\{n_{ij} = 0,1\}} \left(\prod_{i=1}^{2N} \delta\left(1 - \sum_{j=1}^{2N} n_{ij} \right) \right) e^{-\beta E(\{ n_{ij} \})}.$$ (3)

Using integral representations of the δ functions and summing over the n_{ij}'s, this can be conveniently written as :

$$Z = \int_0^{2\pi} \frac{d\lambda_1}{2\pi} \cdots \int_0^{2\pi} \frac{d\lambda_{2N}}{2\pi} \exp\left(i \sum_{j=1}^{2N} \lambda_j \right) \prod_{j<k} [1 + e^{-\beta l_{jk}} e^{-i(\lambda_j + \lambda_k)}].$$ (4)

The basic steps which relate this partition function to the ground state properties of the system have been studied in detail in [8] for the travelling salesman problem. They hold similarly in this case and we shall simply quote the most important facts. The interesting regime of temperature is the low temperature one in which β grows with N in such a way as to pick up in (4) configurations involving short links. In this case the only relevant property of the length distribution $\rho(l)$ is its behaviour around $l = 0$. When :

$$\rho(l) \overset{l \to 0}{\sim} l^r$$ (5)

the nearest neighbour of a given point i is at a distance of order $N^{-1/(r+1)}$ and one expects the energy to scale as $N^{1-1/(r+1)}$. Indeed, this regime is obtained when β scales as :

$$\beta = \hat{\beta} N^{1/(r+1)}$$ (6)

and in that case both the average energy E and the free energy $F = - \dfrac{\partial}{\partial \beta} \log Z$ behave as :

$$E = N^{1-1/(r+1)} \hat{E}; \quad F = N^{1-1/(r+1)} \hat{F}.$$ (7)

One looks for the ground state energy which is the limit of E and F in the limit where $\hat{T} = 1/\hat{\beta} \to 0$.

Clearly, Z in (4), and hence F, still depend on the sample (i.e. of the instance of the l_{ij} one considers). From our experience in the statistical physics of disordered systems we expect however that the free energy in the regime (6) is « self averaging », which means that $\lim_{N\to\infty} F$ exists and is independent of the sample. (In fact if the replica method we use in the following can at all be applied to this problem, it can be used to prove this property so that our method is consistent with respect to the assumption of self averageness). In this case one can deduce F from the value of $\overline{\log Z}$ (in the following ( ) denotes an average with respect to the realizations of the $\{ l_{ij} \}$).

We shall compute $\overline{\log Z}$ with the replica method which consists in computing first $\overline{Z^n}$, from which $\log Z$ is deduced as :

$$\overline{\log Z} = \lim_{n\to 0} \frac{\overline{Z^n} - 1}{n} \tag{8}$$

Z^n is simply the partition function of n uninteracting replicas of the initial system, which is expressed from (4) by introducing on each point i n replicas λ_i^a of the variable λ_i, $a = 1, ..., n$. Then the averages over the distribution of l_{jk} on each link factorize and give (under the conditions (5), (6)) :

$$\int \rho(l_{jk})\, dl_{jk} \prod_{a=1}^{n} [1 + e^{-\beta l_{jk}} e^{-i(\lambda_j^a + \lambda_k^a)}] = 1 + \frac{1}{N} \sum_{p=1}^{n} g_p \sum_{1 \leqslant a_1 < ... < a_p \leqslant n} \exp\left(-i \sum_{r=1}^{p} (\lambda_j^{a_r} + \lambda_k^{a_r}) \right) \tag{9}$$

where :

$$g_p = (p\hat{\beta})^{-(r+1)}. \tag{10}$$

Using well known properties of the Gaussian integrals, $\overline{Z^n}$ can finally be written as :

$$\overline{Z^n} \underset{N\to\infty}{=} \int \prod_{p=1}^{n} \prod_{a_1 < ... < a_p} \frac{dQ_{a_1...a_p}}{\sqrt{2\pi N g_p}} \exp N\left\{ -\frac{1}{2} \sum_{p=1}^{n} \frac{1}{g_p} \sum_{a_1 < ... < a_p} (Q_{a_1...a_p})^2 + 2 \log z \right\} \tag{11}$$

where z is a « one site partition function » :

$$z = \int_0^{2\pi} \frac{d\lambda'}{2\pi} \cdots \int_0^{2\pi} \frac{d\lambda^n}{2\pi} \exp\left(i \sum_{a=1}^{n} \lambda^a + \sum_{p=1}^{n} \sum_{a_1 < ... < a_p} Q_{a_1...a_p} e^{-i(\lambda_{a_1} + ... + \lambda_{a_p})} \right). \tag{12}$$

From (11) the free energy F can be obtained in the limit $N \to \infty$ as the result of a saddle point method. Thanks to the introduction of replicas, the sites have decoupled. This is a nice property of these random-link models that they are naturally mean field theories. As in the case of spin glasses, the whole difficulty consists in finding the solution of the saddle point equations :

$$Q_{a_1...a_p} = 2 g_p \lang\!\langle e^{-i(\lambda_{a_1} + ... + \lambda_{a_p})} \rangle\!\rangle \tag{13}$$

where $\lang\!\langle \cdots \rangle\!\rangle$ denotes one site average, with respect to the one site measure defined in (12). This task is in fact harder than in spin glasses since there all the tensors $Q_{a_1...a_p}$, $1 \leqslant p \leqslant n$ (and not only Q_{ab}) appear as natural order parameters.

We have looked for solutions of (13) in a very restricted space, the replica symmetric one in which :

$$Q_{a_1...a_p} = Q_p. \tag{14}$$

In order to be able to do the continuation to $n = 0$, one must consider cases in which Q_p depends analytically on p. The quantity one has to extremize in (11) is then :

$$-\frac{1}{2} \sum_{p=1}^{n} \frac{1}{g_p} \frac{n !}{p ! (n - p) !} Q_p^2 + 2 \log \left[\left(\frac{\partial}{\partial x} \right)^n \exp \left(\sum_{p=1}^{\infty} \frac{x^p Q_p}{p !} \right) \right]\Big|_{x=0} \right] \tag{15}$$

which in the limit $n \to 0$ is equal to :

$$n \left\{ -\frac{1}{2} \sum_{p=1}^{\infty} \frac{1}{g_p} (- 1)^{p-1} \frac{Q_p^2}{p} + 2 \int_{-\infty}^{+\infty} \mathrm{d}l \, [\mathrm{e}^{-\mathrm{e}^l} - \mathrm{e}^{-G(l)}] \right\} \tag{16}$$

$$G(l) \equiv \sum_{p=1}^{\infty} (- 1)^{p-1} \frac{Q_p \, \mathrm{e}^{l_p}}{p !}. \tag{17}$$

The saddle point equations can be written as a functional equation for $G(l)$. The solution $G_r(l)$ (r characterizes the distribution of distances (5)) satisfies :

$$G_r(l) = \frac{2}{\hat{\beta}^{r+1}} \int_{-\infty}^{+\infty} \mathrm{d}y \, B_r(l + y) \, \mathrm{e}^{-G_r(y)} \tag{18}$$

$$B_r(x) \equiv \sum_{p=1}^{\infty} (- 1)^{p-1} \frac{\mathrm{e}^{xp}}{p^r (p !)^2}. \tag{19}$$

In terms of this solution the thermodynamic quantities are expressed, from (11) and (16), as :

$$\hat{E}_r = \frac{r + 1}{\hat{\beta}} \int_{-\infty}^{+\infty} \mathrm{d}l \, G_r(l) \, \mathrm{e}^{-G_r(l)} \tag{20}$$

$$\hat{F}_r = \frac{\hat{E}_r}{(r + 1)} - \frac{2}{\hat{\beta}} \int_{-\infty}^{+\infty} \mathrm{d}l \, [\mathrm{e}^{-\mathrm{e}^l} - \mathrm{e}^{-G_r(l)}]. \tag{21}$$

The limit of zero temperature ($\hat{\beta} \to \infty$) can be directly studied by the introduction of the function $\hat{G}_r(l) = G_r(\hat{\beta}l)$ which verifies :

$$\hat{G}_r(l) = 2 \int_{-l}^{\infty} \mathrm{d}y \, \frac{(l + y)^r}{r !} \, \mathrm{e}^{-\hat{G}_r(y)} \tag{22}$$

$$\hat{E}_r(\hat{T} = 0) = (r + 1) \int_{-\infty}^{+\infty} \mathrm{d}l \, \hat{G}_r(l) \, \mathrm{e}^{-\hat{G}_r(l)}. \tag{23}$$

Equations (18), (22) are easily solved numerically by iteration for any r. The resulting curve of the free energy *versus* temperature, computed for $r = 1$, exhibits the following typical properties : for finite r there is no phase transition but an important freezing phenomenon; the entropy remains nonnegative and goes to zero with T. The ground state energy is :

$$\hat{E}_{r=1}(\hat{T} = 0) = 1.144. \tag{24}$$

The case $r = 0$ is also interesting. It corresponds to a nonvanishing probability of having infinitely short links. The upper bound on the minimal length from the greedy algorithm for instance is of order $\log N$ [8], while the lower bounds are finite when $N \to \infty$. With our Ansatz,

equations (18) to (23) can be solved for $r = 0$, although one should notice that the corresponding saddle point values of the intermediate variables Q_p are divergent for p large enough, since $G_0(y)$ grows only linearly with y at $y \to \infty$. Nevertheless E and F depend only on G_0 and are well defined. Furthermore the solution of (22) is exactly :

$$\hat{G}_0(l) = \log(1 + e^{2l}) \tag{25}$$

which gives for the ground state energy :

$$\hat{E}_{r=0}(\hat{T} = 0) = \frac{\pi^2}{12}. \tag{26}$$

In order to test the predictions (24) and (26), we have generated the ground states numerically for samples of size $(2\,N =) 20$, 30 and 60 in the case $r = 1$, and 20, 30, 44 for $r = 0$. The procedure consisted in weakening the constraints from $n_{ij} = 0, 1$ to $n_{ij} \in [0, 1]$ and using a linear programming algorithm. Although this sometimes generates (in less than 30 % of the cases for the values of N we consider) spurious states with odd loops of $n_{ij} = 1/2$, it seems that the energies of these spurious states coincide with those of the true ground states in the large N limit. The results we have obtained for the samples in which our method does not give a spurious ground state are plotted in figure 1. The extrapolation to $N \to \infty$:

$$\hat{E}_{r=1}(\hat{T} = 0) = 1.14 \pm 0.01 ; \qquad \hat{E}_{r=0}(\hat{T} = 0) = 0.825 \pm 0.01 \tag{27}$$

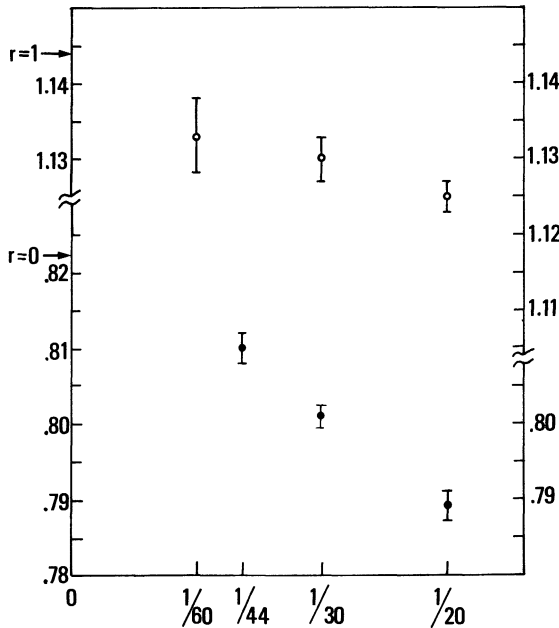are in good agreement with the predicted values.

Fig. 1. — The minimal length (times $N^{-r/(r+1)}$) versus the inverse of the number of points $(1/2\,N)$, for the cases $r = 0$ (full points, lower scale) and $r = 1$ (open points, upper scale). The predicted values are indicated by the arrows. The number of samples studied is, in the case $r = 0$, $16\,000\,(2\,N = 20, 30)$ and $3\,800\,(2\,N = 44)$; and in the case $r = 1$, $7\,000\,(2\,N = 20)$, $1\,800\,(2\,N = 30)$ and $330\,(2\,N = 60)$.

As we have not tested the stability of our solution (i.e. the fact that all the eigenvalues of the Hessian in (11) are nonnegative (^{1})) we cannot be sure that the Ansatz (14) is correct. Some encouraging indications are the fact that the entropy remains positive, and the agreement of the ground state energies with the results of the simulation. A more stringent test comes from the distribution of the lengths of occupied links in the ground state :

$$P_r(L) = \frac{1}{N} \sum_{i<j} \overline{\langle n_{ij} \, \delta(l_{ij} \, N^{1/(r+1)} - L) \rangle}.$$

(28)

This is an interesting information which can be computed with the replica method as follows : we compute the mth moment A_m of $P_r(L)$ by introducing in the partition function an extra weight on each configuration of the type :

$$\exp\left[- \alpha N^{-1+m/(r+1)} \sum_{j<k} n_{jk}(l_{jk})^m \right].$$

(29)

The effect of this weight turns out to change the factors g_p in (10) into :

$$g_p(\alpha) = (p\hat{\beta})^{-(r+1)}\left[1 - \alpha \frac{(r+m)\,!}{r\,!} \frac{p}{N} (p\hat{\beta})^{-m} \right].$$

(30)

Deriving the modified free energy with respect to α, one gets :

$$A_m = \lim_{n\to 0} \frac{1}{2n} \sum_{p=1}^{n} (p\hat{\beta})^{r+1-m} \frac{(r+m)\,!}{r\,!} p \sum_{a_1<...<a_p} (Q_{a_1...a_p})^2$$

(31)

within the Ansatz (14), one can invert these moments explicitly to obtain $P_r(L)$. The result at zero temperature is :

$$P_r(L) = \frac{2}{r\,!} L_r \int_{-\infty}^{+\infty} dl \, \frac{\partial \hat{G}_r}{\partial l} e^{-\hat{G}_r(l) - \hat{G}_r(L-l)}.$$

(32)

We have tested this prediction numerically in the case $r = 0$, where the distribution of the lengths of occupied links (rescaled by a factor N, see (28)) is predicted from (25) and (32) to be :

$$P_{r=0}(L) = \frac{L - e^{-L} \sinh L}{\sinh^2 L}.$$

(33)

The predicted histogram and the numerical results are shown in table I and seem to agree.

An intriguing feature of this solution is that, although replica symmetric, the order parameter turns out to be a whole function $G_r(l)$, and one can wonder whether there is a breaking of ergodicity in this problem as in spin glasses [9]. To answer this question, as in (9), we introduce the

---

(^{1}) In this case, the number of variables in each tensor $Q_{a_1 \, ... \, a_p}$, $n(n - 1) ... (n - p + 1)$, has the sign $(- 1)^{p-1}$ for $n < 1$. Hence the condition of positivity of the eigenvalues of the Hessian corresponds neither to a pure maximum nor to a minimum of (16) in the subspace (17), contrarily to the spin glass case.

Table I. — *Integrated probability distribution of the lengths of the links (multiplied by N) in the optimal configuration in the case r = 0. The first column is the range of integration, the last one is the theoretical prediction obtained from formula* (33). *The average was done on the same number of samples as for the computation of the minimal length (see Fig. 1). The numbers between parentheses are the errors on the last digit.*

Interval	N=20	N=30	N=44	Theory
0.0/ 0.2	.1909 (11)	.1901 (8)	.1895 (15)	.1867
0.2/ 0.4	.1637 (10)	.1626 (8)	.1593 (14)	.1605
0.4/ 0.6	.1390 (9)	.1374 (7)	.1364 (13)	.1356
0.6/ 0.8	.1148 (8)	.1130 (7)	.1141 (12)	.1125
0.8/ 1.0	.0928 (8)	.0919 (6)	.0921 (11)	.0917
1.0/ 1.2	.0736 (7)	.0739 (5)	.0739 (9)	.0736
1.2/ 1.4	.0575 (6)	.0583 (5)	.0581 (9)	.0582
1.4/ 1.6	.0439 (5)	.0450 (4)	.0459 (7)	.0453
1.6/ 1.8⁻	.0342 (5)	.0349 (4)	.0346 (6)	.0349
1.8/ 2.0	.0259 (4)	.0257 (3)	.0264 (6)	.0265
2.0/ 2.5	.0370 (5)	.0382 (4)	.0386 (7)	.0407
2.5/ 3.0	.0163 (3)	.0171 (3)	.0182 (5)	.0190
3.0/ 4.0	.0088 (2)	.0100 (2)	.0110 (4)	.0122
4.0/ 5.0	.00117 (8)	.00153 (7)	.00167 (5)	.00223
5.0/ ∞	.00010 (2)	.00018 (3)	.00034 (2)	.00045

distribution of overlaps between two real copies of the system which have the same $\{ l_{ij} \}$ but different occupation numbers of the links $\{ n_{ij} \}$ and $\{ n'_{ij} \}$ [2] :

$$P(q) = \overline{\left\langle \delta\left(q - \frac{1}{N}\sum_{i<j} n_{ij}\, n'_{ij}\right) \right\rangle_{(2)}} \tag{34}$$

where $\langle \ \rangle_{(2)}$ is a thermal average with a Boltzmann weight $e^{-\beta[E(\{n_{ij}\}) + E(\{n'_{ij}\})]}$. The characteristic function of $P(q)$ can be computed with the replica method [9]; the final result is :

$$P(q) = \lim_{n \to 0} \frac{1}{n(n-1)} \sum_{a \neq b} \delta\left(q - \frac{1}{2}\left\{\frac{1}{g_2}\, Q_{ab}^2 + \sum_{p=1}^{\infty} \frac{1}{g_{p+2}} \sum_{c_1 < \ldots < c_p} (Q_{abc_1\ldots c_p})^2\right.\right). \tag{35}$$

---

[2] Generalized $P(q)$ functions involving the overlap between three $\left((1/N) \cdot \sum_{i<j} n_{ij}\, n'_{ij}\, n''_{ij}\right)$ or more systems can be introduced and computed in a similar way.

A non-trivial $P(q)$ function can be obtained only with replica symmetry breaking. For Ansatz (14), we get :

$$P(q) = \delta(q - q_0); \qquad q_0 = 1 - \int_{-\infty}^{+\infty} dl \ \frac{\partial^2 G_r}{\partial l^2} \ e^{-G_r(l)} \qquad (36)$$

q_0 is an increasing function of the temperature which goes to one when $\hat{T} \to 0$. Clearly an accurate measurement of $P(q)$ is needed in order to test (36) and to establish or infirm the validity of the replica symmetric Ansatz.

To conclude we would like to make a comment on the generality of the above approach. Although we have described the method in the special case of the weighted matching, we think it can be applied to other problems as well. An interesting variant is the bipartite weighted matching problem [7], which we have solved in the same way. Under the assumption of the symmetry between the two types of points, in this case we recover formulas (18) to (23) exactly, except for the explicit factors 2 in (18) and (22) which should be replaced by 1. It follows that the ground state energy in the bipartite case is $2^{1/(r+1)}$ larger than in the problem we have studied. This bipartite matching is also interesting since the linear programming algorithm does not give any spurious solution. The replica method can also be applied to NP complete problems, and in these cases it will be interesting to see whether one needs replica symmetry breaking.

References

[1] KIRKPATRICK, S., *Lecture Notes in Physics* Vol. **149** (Springer, Berlin) 1981, p. 280 ;
    KIRKPATRICK, S., GELATT, C. D. Jr., VECCHI, M. P., *Science* **220** (1983) 671.
[2] For a review, see e.g. : « Heidelberg colloquium on Spin Glasses », *Lecture Notes in Physics* Vol. **192** (Springer, Berlin) 1983 ;
    PARISI, G., lectures given at the 1982 Les Houches summer school (North Holland).
[3] BARAHONA, F., *J. Phys. A* **15** (1982) 3241.
[4] SIARRY, P., DREYFUS, G., *J. Physique Lett.* **45** (1984) L-39.
[5] KIRKPATRICK, S., TOULOUSE, G., preprint (1985).
[6] BACHAS, C. P., *Phys. Rev. Lett.* **54** (1985) 53.
[7] PAPADIMITRIOU, C. H., STEIGLITZ, K., *Combinatorial Optimization* (Prentice Hall) 1982.
[8] VANNIMENUS, J., MÉZARD, M., *J. Physique Lett.* **45** (1984) L-1145.
[9] PARISI, G., *Phys. Rev. Lett.* **50** (1983) 1946.