Mean Field and Corrections for the Euclidean Minimum Matching Problem

Jacques H. Boutet de Monvel* and Olivier C. Martin†

Division de Physique Théorique, Institut de Physique Nucléaire, Université Paris-Sud, F-91406 Orsay Cedex, France

(Received 24 January 1997)

The minimum matching of N random points in d-dimensional Euclidean space is a tractable model of frustration with disorder. We use numerical simulations to obtain precise estimates of the ground-state energy for $2 \leq d \leq 10$. We then consider the approximation where distance correlations are neglected. This model's solution leads to an excellent "random link" approximation at $d \geq 2$. Incorporating three-link correlations improves the accuracy, leading to a relative error of 0.4% at $d = 2$ and 3. Finally, we argue that the Euclidean model's $1/d$ series is beyond all orders of a link correlation expansion. [S0031-9007(97)03532-1]

PACS numbers: 75.50.Lk, 64.60.Cn

There has been a tremendous amount of work on frustrated disordered systems in the past twenty years, in part driven by the exact solution [1] of the Sherrington-Kirkpatrick spin glass model [2]. However, the relevance of this infinite range model to three-dimensional short range systems remains controversial [3,4]. Two of the difficulties which have hampered attempts [5,6] to use replicas in finite dimensions are: (i) for finite connectivity systems, the saddle point equations involve the order parameters $q_{\alpha\beta\gamma\cdots}$ with an arbitrary number of indices rather than just $q_{\alpha\beta}$ [7]; (ii) in d-dimensional Euclidean space, the geometry introduces strong correlations among the quenched bond variables because it selects which bonds are present. It is usually overlooked that these two obstacles have been overcome in a frustrated disordered system simpler than a spin glass, the minimum matching problem (MMP) [8]. For this problem, Mézard and Parisi have treated the order parameters $q_{\alpha\beta\gamma\cdots}$ exactly [9], and have introduced an expansion [10] to take into account the correlations among the bond variables. In this Letter, we determine the accuracy of their replica approach by comparing with the actual properties of the d-dimensional Euclidean MMP. First, we find that the relative error introduced by the leading approximation for the zero-temperature energy density is less than 4% at $d = 2$ and 3% at $d = 3$. Second, the inclusion of the first Euclidean corrections reduces that error by a factor of about 10 at $d = 2$ and $d = 3$, showing that the replica approach gives a quantitatively accurate description of this system in finite and even low dimensions. Third, we argue that the large d behavior of systems such as the MMP and possibly spin glasses depends on arbitrarily high order bond correlations and is thus beyond all orders of the expansion proposed by Mézard and Parisi [10].

Consider N points (N even) and a specified set of link lengths $l_{ij} = l_{ji}$ separating the points, for $1 \leq i, j \leq N$. One defines a matching (a dimerization) of these points by combining them pairwise so that each point belongs to one and only one pair. Define also the energy or length of a matching as the sum of the lengths of the links associated with each matched pair. The minimum matching problem is the problem of finding the matching of minimum energy. The constraint of having a dimerization introduces frustration as in general all points cannot be paired to their nearest neighbor. One can also consider the thermodynamics of this system, as proposed by Orland [11] and Mézard and Parisi [12], by taking *all* matchings but weighting them with the Boltzmann factor. Here we concentrate on the $T = 0$ properties because exact ground states can be obtained for any given instance of size N using on the order of N^3 steps [13].

Physically, one is not interested in the properties of any particular instance of the MMP; more relevant are typical and ensemble properties such as the average energy when the lengths l_{ij} are random variables with a given distribution. One then speaks of the stochastic MMP. There are two frequently used ensembles for the l_{ij}, corresponding to the *Euclidean* MMP and the *random link* MMP. In the first, the N random points lie in a d-dimensional Euclidean volume and the l_{ij} are the usual Euclidean distances between pairs of points. The points are independent and identically distributed, so one speaks of a random point problem. In the second ensemble, it is the link lengths l_{ij} which are independent and identically distributed random variables. A connection between these two systems was first given by Mézard and Parisi [9]: they pointed out that the one- and two-link distributions could be made identical in the two problems. A consequence is that the Cayley tree approximation for the random point and random link problems are the same. Mézard and Parisi were able to solve the random link MMP using an approach based on replicas [9,12]. One may then consider the random link MMP to be a "mean field model" for the Euclidean MMP. The mean field *approximation* consists of using the thermodynamic functions of the random link model as estimators for those of the Euclidean model. Hereafter we shall refer to it as the *random link approximation* [14]. Finally, Mézard and Parisi have shown how to derive corrections to the random link approximation using an expansion in link correlations. In [10], they have computed the leading corrections, associated with the triangle inequality in the Euclidean model.

How accurate are these approximations? To answer this, we first give our numerical results for the Euclidean MMP, and then compare with these analytical predictions.

In the Euclidean MMP, let L^E_{MM} be the energy or length of the minimum matching. Taking the points to be independent and uniformly distributed in a unit volume, Steele [15] has shown that as $N \to \infty$, $L^E_{MM}/N^{1-1/d}$ converges with probability one to a nonrandom, N-independent constant $\beta^E_{MM}(d)$. In physics language, this result shows that L^E_{MM} is self-averaging and that the zero-temperature energy density has an infinite volume limit when the density of points is kept fixed. To date, little has been done to compute the ground-state energy densities $\beta^E_{MM}(d)$. The best estimates are $\beta^E_{MM}(2) \approx 0.312$ [16,17] and $\beta^E_{MM}(3) \approx 0.318$ [16]. Here we use a systematic procedure [18] to obtain $\beta^E_{MM}(d)$ with quantifiable errors. First, in order to have a well defined dependence on N, we have used the ensemble average, $\langle L^E_{MM}\rangle/N^{1-1/d}$. Second, in order to reduce corrections to scaling in the extrapolation to the large N limit, we have placed the points randomly in the d-dimensional unit hypercube with periodic boundary conditions. This removes surface effects and empirically leads to the finite size scaling law

$$\frac{\langle L^E_{MM}\rangle}{N^{1-1/d}} = \beta^E_{MM}(d)\left(1 + \frac{A(d)}{N} + \frac{B(d)}{N^2} + \ldots\right). \quad (1)$$

Finally, in order to reduce statistical fluctuations,we have used a variance reduction trick [14]. Using these methods, we have extracted from our numerical data $\beta^E_{MM}(d)$ and its associated statistical error. The fits to Eq. (1) are good, with χ^2 values confirming the form of the finite size scaling law. The error bars on the extrapolated value $\beta^E_{MM}(d)$ are obtained by requiring that χ^2 increase by one from its minimum. We find $\beta^E_{MM}(2) = 0.3104 \pm 0.0002$, and $\beta^E_{MM}(3) = 0.3172 \pm 0.000\,15$; values at higher dimensions are given in Table I.

Now we discuss how to use the random link model to approximate $\beta^E_{MM}(d)$. For any two points (i, j) placed at random in the unit d-dimensional hypercube, the density distribution of l_{ij} is given at short distances by $P_d(l_{ij} = r) = dB_d r^{d-1}$, where $B_d = \pi^{d/2}/(d/2)!$ is the volume of the d-dimensional ball with unit radius. If we take the random link model where link lengths are independent and have the individual distribution $P_d(l)$, then the Euclidean and random link MMP have the same one- *and* two-link distributions [9] because two Euclidean distances are independent. If correlations among three or more link lengths are weak, then the properties of the two systems should be quantitatively close. Thus an analytic approximation to $\beta^E_{MM}(d)$ is obtained by computing its analog $\beta^{RL}_{MM}(d)$ in the random link MMP. In Refs. [9,12], Mézard and Parisi solved these random link models under the replica symmetry hypothesis. They showed further that the replica symmetric solution is stable (at least for $d = 1$), and thus is most likely exact. Their solution gives $\beta^{RL}_{MM}(d)$ in terms of a function G_d related to the probability distribution of link lengths for matched pairs. In our Euclidean units their result can be written

$$\beta^{RL}_{MM}(d) = \frac{D_1(d)}{2}\,\frac{d}{(1/d)!}\int_{-\infty}^{+\infty} G_d(x)e^{-G_d(x)}\,dx\,, \quad (2)$$

where G_d satisfies the integral equation

$$G_d(x) = d\int_{-x}^{+\infty}(x + y)^{d-1}e^{-G_d(y)}\,dy\,, \quad (3)$$

and where

$$D_1(d) = \lim_{N\to\infty}\langle L_1\rangle/N^{1-1/d} = (1/d)!\,B_d^{-1/d} \quad (4)$$

is the average (rescaled) link length of the nearest neighbor graph in the limit $N \to \infty$.

Brunetti *et al.* [19] have used direct numerical simulations of these random link models to confirm the predictions to the level of 0.2% at $d = 1$ and 2, and we have done the same to the level of 0.1% at $1 \le d \le 10$, giving further evidence that the replica symmetric solution is exact. From the analytical side, solving the integral equation for G_d leads to $\beta^{RL}_{MM}(1) = \pi^2/24 = 0.411\,233\,5\ldots$, $\beta^{RL}_{MM}(2) = 0.322\,580\ldots$, and $\beta^{RL}_{MM}(3) = 0.326\,839\ldots$; values at higher dimensions are given in

TABLE I. Comparison of MMP ground-state energies for the three models: Euclidean, random link, and random link including three-link Euclidean corrections ($2 \le d \le 10$).

d	$\beta^E(d)$	$\beta^{RL}(d)$	$d\frac{\beta^{RL}-\beta^E}{\beta^E}$	$\beta^{EC}(d)$	$\frac{\beta^{EC}-\beta^E}{\beta^E}$
2	0.3104 ± 0.0002	0.322 580	+0.078	0.309 15	−0.40%
3	0.3172 ± 0.000 15	0.326 839	+0.091	0.318 26	+0.33%
4	0.3365 ± 0.0003	0.343 227	+0.080	0.337 56	+0.30%
5	0.3572 ± 0.000 15	0.362 175	+0.070	0.358 18	+0.27%
6	0.3777 ± 0.0001	0.381 417	+0.059	0.378 49	+0.21%
7	0.3972 ± 0.0001	0.400 277	+0.054	0.398 07	+0.22%
8	0.4162 ± 0.0001	0.418 548	+0.045	0.416 85	+0.17%
9	0.4341 ± 0.0001	0.436 185	+0.042	0.434 85	+0.17%
10	0.4515 ± 0.0001	0.453 200	+0.037	0.452 14	+0.14%

Table I. If we consider $\beta_{MM}^{RL}(d)$ as a mean field prediction for $\beta_{MM}^{E}(d)$, the accuracy is surprisingly good. Including the trivial value $\beta_{MM}^{E}(1) = 0.5$, we see that the random link approximation leads to a relative error of 17.8% at $d = 1$, of 3.9% at $d = 2$, and of 3.0% at $d = 3$. Also, the error decreases with increasing dimension. It can be argued, for the MMP as well as for other link-based problems [14], that the random link approximation not only has a relative error tending towards 0 as $d \to \infty$, but that, in fact, this error is at most of order $1/d^2$. Given our high quality estimates, we are able to confirm this property numerically. In Fig. 1 we plot the quantity $d(\beta_{MM}^{RL} - \beta_{MM}^{E})/\beta_{MM}^{E}$ along with a quadratic fit given to guide the eye. As expected, the data scales as $1/d$. Thus the random link approximation gives both the leading and $1/d$ subleading dependence of $\beta_{MM}^{E}(d)$. In order to obtain analytic expressions for the associated coefficients, we have derived the $1/d$ expansion for β_{MM}^{RL} from Eqs. (2) and (3). We used two methods to do this. The first, straightforward but computationally lengthy, consists of setting $\tilde{G}_d(x) = G_d(\tilde{x} = x/d + 1/2)$ and then writing $\tilde{G}_d(x)$ as a power series in $1/d$. From this we find

$$\beta_{MM}^{RL}(d) = \frac{D_1(d)}{2}\left[1 + \frac{1-\gamma}{d} + O(1/d^2)\right], \qquad (5)$$

where $\gamma = 0.577\ldots$ is Euler's constant. If, as claimed, the random link approximation gives an error of order $1/d^2$, Eq. (5) gives an analytic expression for the leading and first subleading terms in the $1/d$ expansion of $\beta_{MM}^{E}(d)$. This claim is strongly supported by the numerical results: performing a fit of our $\beta_{MM}^{E}(d)$ values to a truncated $1/d$ series leads to 0.424 ± 0.008 for the coefficient of the $1/d$ term; this is to be compared to the theoretical prediction of $1 - \gamma = 0.422\,784\ldots$.

We have been able to obtain the next coefficient of the series in $1/d$ for β_{MM}^{RL} by using a second method. We introduce a modified random link model where the links

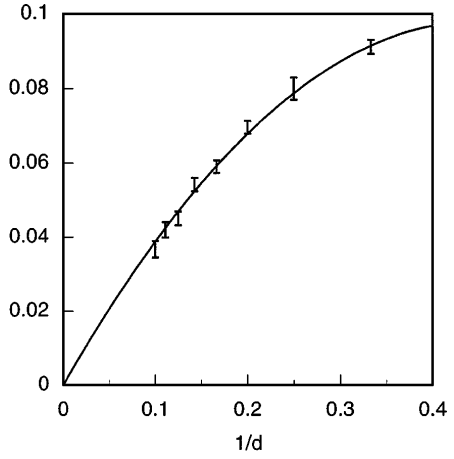

FIG. 1. Linear scaling with $1/d$ of the quantity $d(\beta_{MM}^{RL} - \beta_{MM}^{E})/\beta_{MM}^{E}$.

are shifted and rescaled in such a way that the leading term of the $1/d$ expansion for this new model is exactly the $1/d$ coefficient for the initial one [20]. In fact, it is possible to introduce a sequence of such "rescaled" models, where the kth model is designed to produce the $1/d^k$ term of the expansion. We have computed the leading terms predicted by a replica symmetric analysis of these models for $k = 1$ and 2, from which we find that the order $1/d^2$ coefficient in Eq. (5) is $\pi^2/12 + \gamma^2/2 - \gamma$.

We now come to the final point of the paper: how well can one predict $\beta_{MM}^{E}(d)$ by incorporating Euclidean corrections to the random link approximation? It is necessary here to review the work of Mézard and Parisi; for greater detail, we refer the reader to their article [10]. They begin with the partition function Z for an arbitrary stochastic MMP and write the quenched average for n replicas. In the Euclidean model, the l_{ij} have three- and higher-link correlations. Mézard and Parisi keep the three-link correlations (arising only when the three links make a triangle) and neglect higher connected correlations. Note that it is not clear *a priori* whether these "higher order" terms are negligible compared to the three-link term. The resulting expression for the quenched average becomes

$$\overline{Z^n} = \prod_{j=1}^{N}\prod_{\alpha=1}^{n}\left(\int_0^{2\pi}\frac{d\lambda_j^{\alpha}}{2\pi}\,e^{i\lambda_j^{\alpha}}\right) \times e^{\sum_{(ij)}\overline{u_{ij}} + \sum'_{(ij)(kl)(mn)}\overline{u_{ij}u_{kl}u_{mn}}^{C}}, \qquad (6)$$

where u_{ij} is a complicated nonlinear function of the link length l_{ij}. They then compute the limit $N \to \infty$, $n \to 0$ using the saddle point method while assuming that replica symmetry is not broken. In the zero-temperature limit, just as in the standard random link model, the saddle point equations can be expressed in terms of G_d, but now G_d satisfies a more complicated integral equation [Eq. (34) in their paper]. From this, one can calculate new estimates for $\beta_{MM}^{E}(d)$, which we shall denote β_{MM}^{EC}, where EC stands for Euclidean corrections.

We have solved the equations numerically for this modified G_d, and have computed $\beta_{MM}^{EC}(d)$ for $2 \le d \le 10$. We find $\beta_{MM}^{EC}(2) = 0.30915$ and $\beta_{MM}^{EC}(3) = 0.318\,26$. The results for $d \ge 4$ are given in Table I. Comparing with $\beta_{MM}^{E}(d)$ and $\beta_{MM}^{RL}(d)$, we see that the new estimates are considerably more accurate. At $d = 2$, the random link approximation leads to an error of 3.9%; this error is decreased by nearly a factor of 10 by incorporating these leading Euclidean corrections. Similarly at $d = 3$, the error is reduced from 3.0% to less than 0.4%. At larger d, the error continues to decrease, but the effect is less dramatic.

To interpret this last result, consider how the difference $\beta_{MM}^{EC} - \beta_{MM}^{RL}$ scales with d. Using Eq. (6), we see that it is sufficient to estimate the size of the three-link correction term. Its d dependence follows that of the probability of finding nearly equilateral triangles as $d \to \infty$. Since this

probability goes to zero exponentially with d, the three-link correlations give tiny corrections at large d (as confirmed by the numerics), and the power series expansion in $1/d$ of β_{MM}^{EC} is *identical* to that of β_{MM}^{RL}. This property continues to hold if one includes 4, 5, or any *finite* number of multilink correlations in Eq. (6). This is due to the fact that the Euclidean and random link graphs have *local* properties that are identical up to exponentially small terms in d. In particular, the statistics of fixed sized (N-independent) loops connecting near neighbors are nearly identical.

Although this reasoning was given for the MMP, it applies equally well to other link-based problems. In such statistical mechanics systems, if the thermodynamic functions depend only on the local properties of the (short) link graph, then the random link approximation applied to the Euclidean system will have an error which is exponentially small in d. However, for combinatorial optimization problems such as the MMP, the assignment problem, and the traveling salesman problem, the $N \to \infty$ limit and the k-link expansion do not commute: k-link correlations with k growing with N remain important as $N \to \infty$. In particular, arbitrarily large loops matter and contribute to the thermodynamics at order $1/d^2$. In a polymer picture, we can say that the random link approximation is exponentially good in the dilute phase, while it leads to $1/d^2$ errors in the collapsed phase. The $1/d$ power series in this phase is beyond all orders in a k-link correlation expansion such as Eq. (6). This behavior is reminiscent of that of lattice models where the $1/d$ expansion does not commute with the $N \to \infty$ limit (an example is the Eden model [21]). Possibly a similar phenomenon occurs for spin glasses, rendering the calculation of the $1/d$ series particularly difficult.

In summary, we have estimated by numerical simulation $\beta_{MM}^{E}(d)$, the ground-state energy density in the Euclidean minimum matching problem at dimensions $2 \leq d \leq 10$. We have computed two analytical estimates for these energy densities, namely, $\beta_{MM}^{RL}(d)$ and $\beta_{MM}^{EC}(d)$. The first method uses the random link approximation where all link correlations are neglected. Using the exact solution of Mézard and Parisi, we find that even at low dimensions, the error introduced by this approximation is small: 3.9% at $d = 2$, 3.0% at $d = 3$, and 2.0% at $d = 4$. In the second method, the connected three-link correlations are taken into account while higher ones are neglected. Using Mézard and Parisi's expressions, we find that this modification gives excellent predictions even at $d = 2$ and 3, with the error there being divided by almost 10 compared to the random link approximation. This provides a stringent quantitative test of a systematic expansion which goes beyond uncorrelated disorder variables, and suggests that even the leading such correction is enough to get predictions for thermodynamic functions precise to better than 1%. However, at *high* dimensions, we have seen that this k-link correlation expansion cannot give the $1/d$ expansion of the MMP; probably this is not special to the MMP, but applies more generally to most link-based combinatorial optimization problems and perhaps even to spin glasses. That leaves open the analytical determination of the $1/d^2$ coefficient for these kinds of problems.

We are grateful to O. Bohigas, C. De Dominicis, R. Friedberg, J. Houdayer, W. Krauth, M. Mézard, H. Orland, and A. Percus. J. B. dM. acknowledges a fellowship from the MENESR, and O. C. M. acknowledges support from the Institut Universitaire de France. The Division de Physique Théorique is an Unité de Recherche des Universités Paris XI et Paris VI associée au CNRS.

*Electronic address: boutet@ipno.in2p3.fr
†Electronic address: martino@ipno.in2p3.fr

[1] G. Parisi, J. Phys. A **13**, L115 (1980).
[2] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).
[3] D. Fisher and D. Huse, Phys. Rev. B **38**, 386 (1988).
[4] D. Fisher and H. Sompolinsky, Phys. Rev. Lett. **54**, 1063 (1985).
[5] A. Georges, M. Mézard, and J. Yedidia, Phys. Rev. Lett. **64**, 2937 (1990).
[6] C. D. Dominicis, I. Kondor, and T. Temesvari, J. Phys. A **24**, L301 (1991).
[7] C. D. Dominicis and P. Motishaw, Europhys. Lett. **3**, 87 (1987).
[8] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Prentice Hall, Englewood Cliffs, NJ, 1982).
[9] M. Mézard and G. Parisi, Europhys. Lett. **2**, 913 (1986).
[10] M. Mézard and G. Parisi, J. Phys. (Paris) **49**, 2019 (1988).
[11] H. Orland, J. Phys. (Paris), Lett. **46**, L763 (1985).
[12] M. Mézard and G. Parisi, J. Phys. (Paris), Lett. **46**, L771 (1985).
[13] M. O. Ball and U. Derigs, Networks **13**, 517 (1983).
[14] N. J. Cerf *et al.*, J. Phys. I (France) **7**, 117 (1997).
[15] M. Steele, The Annals of Probability **9**, 365 (1981).
[16] W. D. Smith, Ph.D. thesis, Princeton University, Princeton, NJ, 1989.
[17] P. Grassberger and H. Freund, Z. Oper. Res. **34**, 239 (1990).
[18] A. G. Percus and O. C. Martin, Phys. Rev. Lett. **76**, 1188 (1996).
[19] R. Brunetti, W. Krauth, M. Mézard, and G. Parisi, Europhys. Lett. **14**, 295 (1991).
[20] J. Boutet de Monvel, Ph.D. thesis, Institut de Physique Nucléaire, Université Paris-Sud, Orsay, 1996.
[21] R. Friedberg, Phys. Lett. A **112**, 129 (1985).