PHYSICAL REVIEW E **74**, 041122 (2006)

# Statistical mechanics of multi-index matching problems with site disorder

David S. Dean[1] and David Lancaster[2]
[1]*Laboratoire de Physique Théorique, IRSAMC, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 04, France*
[2]*Harrow School of Computer Science, University of Westminster, Harrow, HA1 3TP, United Kingdom*

(Received 3 July 2006; published 25 October 2006)

We study the statistical mechanics of multi-index matching problems where the quenched disorder is a geometric site disorder rather than a link disorder. A recently developed functional formalism is exploited that yields exact results in the finite-temperature thermodynamic limit. Particular attention is paid to the zero-temperature limit of maximal matching problems where the method allows us to obtain the average value of the optimal match and also sheds light on the algorithmic heuristics leading to that optimal match.

DOI: 10.1103/PhysRevE.74.041122 PACS number(s): 05.20.−y, 89.20.−a, 89.75.−k

## I. INTRODUCTION

Matching problems have been studied from both the perspectives of combinatorial optimization [1] and statistical physics [2]. In the statistical physics approach, interest lies in computing average properties of the matching path in stochastic versions of the problem. The most general form of the $m$-partite multi-index problem is stated as follows. One has $m$ sets of $N$ objects and a matching consists of combining one element from each set into a molecule or match with $m$ components. If in the $i$th match the element we choose from set $j$, $1 \leqslant j \leqslant m$, is denoted by $i_j$, then the cost of the $i$th match is given by the entry $C_{i_1,\ldots,i_m}$ of the cost matrix. The total cost of the matching is then given by the sum of the costs of the $N$ individual matches. The underlying maximum (minimum) optimization problem consists of finding the global match that maximizes (minimizes) this cost. The bipartite matching problem with $m=2$ has been most studied, and some analytic progress has been made for so-called "random link" models where the elements of the cost matrix $C$ are chosen independently from some probability distribution [2–4].

Often matching problems have a geometrical origin as the points to be matched lie in some $d$-dimensional domain $\mathcal{D}$. Euclidean matching problems are of practical importance, finding applications in air traffic control [5] and the manufacture of printed circuits [6]. The three-dimensional Euclidean matching problem, both minimal and maximal, has been studied numerically via Monte Carlo simulations and in particular simulated annealing [7]. In these geometric problems the matching matrix $C$ is determined from the location of the points using some metric on the space. As a result matrix elements are no longer independent (for example, triangle inequalities exist for the case when the matching matrix depends on the Euclidean distance between the points). The random link versions can be regarded as approximations to Euclidean models, and in minimal bipartite matching this approximation seems rather good [8] and corrections to account for triangular correlations have been computed [3].

Recently, we have developed additional techniques for analyzing Euclidean traveling-salesman-type problems (TSPs) [9] based on functional integration and a functional order parameter. This formalism does not yield the average value of the optimal path for the minimal TSP because it does not apply when the energy is rescaled to make the ground state energy extensive. However, in the case of maximal TSP-type problems the ground state energy is extensive without any energy rescaling and the average value of the optimal solution can be computed exactly. The method [9] does not rely on the replica trick; however, the results obtained show that for the Euclidean TSP the system is replica symmetric, as is the case for random link calculations [2]. In this paper we show how the formalism of [9] can be applied to multi-index matching problems at finite temperature and also how it can be used to obtain the average value of the optimal match in the maximal case. In addition the method gives us information concerning the algorithmic heuristic leading to the optimal path. The mathematical structure of the multi-index matching problem in our thermodynamic formalism is more complicated than that of the Euclidean TSP in that a functional order parameter appears for each of the $m$ sets of points. However, in the cases studied here the functional order parameters are the same for each set and we find no evidence of symmetry breaking. As a consequence of this symmetry we find that the thermodynamics of the bipartite matching problem and the Euclidean TSP are essentially equivalent. For certain other multi-index problems, with chainlike matching functions, we find relations between the average path lengths computed in these models and the TSP.

The statistical mechanical cavity approach has recently been applied to minimal multi-index matching problems in the context of independent links [4]. In this problem the energy (or equivalently the temperature) is scaled to ensure that the ground state energy, corresponding to the maximal match, is extensive. In contrast to the bipartite matching problem which has a replica symmetric solution, it was shown that the multi-index matching problem for more than two indices has a low-temperature glassy phase characterized by replica symmetry breaking. It is therefore interesting to see if the formalism developed in [9], which is exact for maximal optimization problems, shows similar behavior when applied to multi-index matching. In all of the cases studied here the system exhibits no such phase transition, suggesting that they appear to be nonglassy from the statistical mechanics point of view. We have not been able to carry out the calculation in our formalism with a temperature rescaling, which would give an extensive ground state energy for minimal matching. It is highly probable that the rescaled site disordered problem does exhibit a phase transition, at

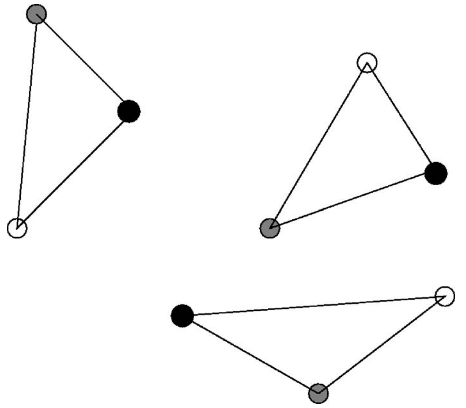

FIG. 1. Full match for three sets (white, black, gray) of three points ($m=3$, $N=3$).

least above some critical dimension. The extension of our functional method to the rescaled temperature regime remains as a difficult and open challenge. We note, however, that the random link matching problem without temperature rescaling does not present any phase transition and this is entirely compatible with our results on site disordered matching in the same regime of scaling.

## II. MODEL AND GENERAL SOLUTION

Here we define the general multi-index matching problem with site disorder.

Consider the following $m$-partite matching problem with $m=K+1$ sets $S^{(0)},S^{(1)},\ldots,S^{(K)}$, each set $S^{(a)}$ consists of $N$ points $\{\mathbf{r}_1^{(a)},\ldots,\mathbf{r}_N^{(a)}\}$ distributed in a common domain $\mathcal{D}$ in a space of dimension $d$. We assume that the points are independently distributed within the domain and that those in set $S^{(a)}$ are distributed according to a probability density $p_a$. An individual match consists of $K+1$ points where one point in each set is matched with one point in each other set. Each match has an energy function associated with it that we denote by $V(\mathbf{r}^{(0)},\mathbf{r}^{(1)},\ldots,\mathbf{r}^{(K)})$. The multi-index optimization problem involves making $N$ individual matches (Fig. 1), where each point is associated with one and only one matching, and optimizing the total match energy. The case of $K=1$ corresponds to the well-known bipartite matching problem. A microstate of the model is specified by $\{(i,\sigma_i^{(1)},\ldots,\sigma_i^{(K)})\}$ where the point $i$ of the set $S^{(0)}$ having position $\mathbf{r}_i^{(0)}$ $(1\leq i\leq N)$ is matched with $\mathbf{r}^{(1)}_{\sigma_i^{(1)}},\ldots,\mathbf{r}^{(K)}_{\sigma_i^{(K)}}$ in the other sets and $\sigma_i^{(a)}$ denotes the label of the element in set $a$ chosen to be matched with the element $i$ in set 0. The $\sigma^{(a)}$ are permutations on the $N$ elements of $S^{(a)}$. The phase space for multi-index matching is thus the Cartesian product $(\Sigma_N)^K$, where $\Sigma_N$ is the permutation group on $N$ elements. The size of this phase space is $(N!)^K$ and the entropy is consequently superextensive.

The Hamiltonian of the system is given by adding up the energy for each individual match

$$H[\sigma_1,\sigma_2,\ldots,\sigma_K]=\sum_{i=1}^{N} V(\mathbf{r}_i^{(0)},\mathbf{r}^{(1)}_{\sigma_i^{(1)}},\ldots,\mathbf{r}^{(K)}_{\sigma_i^{(K)}}), \tag{1}$$

and the partition function for sets of $N$ points is given by

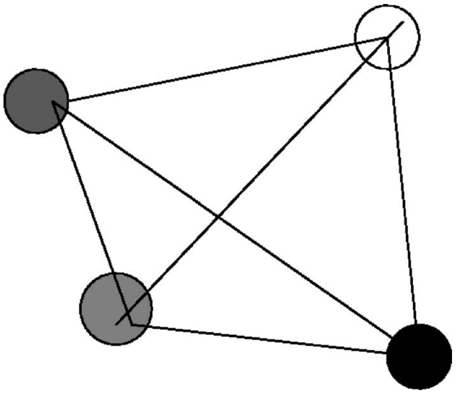

FIG. 2. Individual match between four points (each in a different set white, light gray, dark gray and black), showing a fully symmetric and fully connected cost function.

$$Z_N=\sum_{\sigma^{(1)},\sigma^{(2)},\ldots,\sigma^{(K)}} \exp(-\beta H[\sigma_1,\sigma_2,\ldots,\sigma_K]), \tag{2}$$

where $\beta=1/T$ and $T$ is the canonical temperature of the system.

The form of $V$ can be arbitrary, but a symmetric potential function where $V(\mathbf{r}_0,\ldots,\mathbf{r}_K)$ is left invariant by any rearrangement of its arguments is natural in this context. A simple way of constructing such a potential is to consider the potential to be the totally connected sum of pairwise matching potentials between all pairs as shown in Fig. 2:

$$V(\mathbf{r}_0,\mathbf{r}_1,\ldots,\mathbf{r}_K)=\sum_{a=1,b=1}^{K} V_P(\mathbf{r}_a,\mathbf{r}_b). \tag{3}$$

Practically we see that matching costs of the above type are suited to the identification of cliques or groups.

However, to make analytic progress, much of our analysis considers potentials that are pairwise but are not symmetric. These potentials allow the match to be represented as an ordered path through points from each successive set as illustrated in Fig. 3:

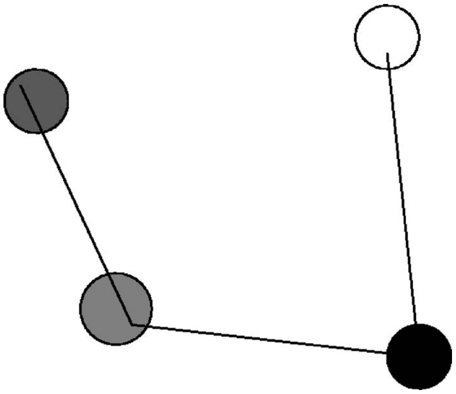

FIG. 3. Individual match between four points (each in a different set white, light gray, dark gray and black), showing an open chain cost function.

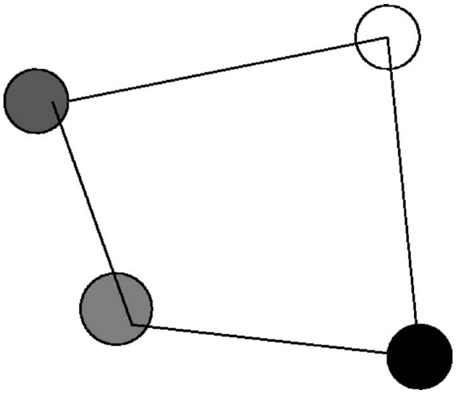

FIG. 4. Individual match between four points (each in a different set white, light gray, dark gray and black), showing a closed chain cost function.

$$V(\mathbf{r}_0,\mathbf{r}_1,\ldots,\mathbf{r}_K)=\sum_{a=1}^{K}V_P(\mathbf{r}_{a-1},\mathbf{r}_a). \tag{4}$$

This type of matching cost is of the type that one would use if trying to reproduce trajectories from a sequence of images of an ensemble of identical objects.

Besides this open model, a generalization is to add $V_P(\mathbf{r}_K,\mathbf{r}_0)$, to arrive at the situation shown in Fig. 4 and we shall term this a closed or cyclic model.

Because we emphasize the geometric interpretation of pairwise matching, we will frequently take the pair potential $V_P$ to be the Euclidean distance $V_P(\mathbf{r}_{i-1},\mathbf{r}_i)=|\mathbf{r}_{i-1}-\mathbf{r}_i|$, though we also treat the quadratic form which seems more amenable to analytic approaches. The case $K=1$ corresponds to the Euclidean bipartite matching problem. The cyclic, $K=2$ tripartite case is symmetric and corresponds to the triangle interaction, which amounts to the total length needed to form a triangle between three points. Another symmetric possibility for $K=2$, tripartite matching, that is not of the pairwise form, is the star interaction where the three points are connected by three lines emanating from the point that would be their center of mass (for equal masses)

$$V_{star}(\mathbf{r}_0,\mathbf{r}_1,\mathbf{r}_2)=|\mathbf{r}_0-\mathbf{R}|+|\mathbf{r}_1-\mathbf{R}|+|\mathbf{r}_2-\mathbf{R}|, \tag{5}$$

where

$$\mathbf{R}=\frac{1}{3}(\mathbf{r}_0+\mathbf{r}_1+\mathbf{r}_2). \tag{6}$$

As pointed out in [9] the disorder in this type of problem can be encoded in the unaveraged density of the quenched points in the domain $\mathcal{D}$

$$\rho_a(\mathbf{r})=\frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{r}-\mathbf{r}_i^{(a)}); \tag{7}$$

note that the value of $\rho_a$ averaged over the disorder is by definition $p_a$.

Imposing this constraint we may discard the sum over permutations and write

$$Z_N=C_N\int\prod_{i,a}d\mathbf{r}_i^{(a)}\prod_{\mathbf{r}}\prod_{a}\delta\left(N\rho_a(\mathbf{r})-\sum_{i=1}^{N}\delta(\mathbf{r}_i^{(a)}-\mathbf{r})\right)\times\exp\left(-\beta\sum_i V(\mathbf{r}_i^{(0)},\ldots,\mathbf{r}_i^{(K)})\right). \tag{8}$$

In the above formulation without the $\delta$ function constraints the matches chosen can use any point in the domain $\mathcal{D}$; however the $\delta$ function constraint allows the use of only the points that are available in the sets $S_a$. The $\delta$ function constraints that are present at each point $\mathbf{r}$ in the domain $\mathcal{D}$ are now expressed as a functional Fourier integral (the $\mu$ integration is along the imaginary axis)

$$Z_N=C_N'\int\prod_a d[\mu_a]\prod_{i,a}d\mathbf{r}_i^{(a)}\exp\left(N\int d\mathbf{r}\sum_a\mu_a(\mathbf{r})\rho_a(\mathbf{r})\right)\times\exp\left(-\sum_{ai}\mu_a(\mathbf{r}_i^{(a)})-\beta\sum_i V(\mathbf{r}_i^{(0)},\ldots,\mathbf{r}_i^{(K)})\right). \tag{9}$$

The integrals over the dynamical variables $\mathbf{r}_i^{(a)}$that are taken to be in $\mathcal{D}$ may now be carried out independently, changing the normalization and leaving an integration over only $K+1$ functional variables:

$$Z_N=C_N''\int\prod_a d[\mu_a]\exp\left(N\left\{\int d\mathbf{r}\sum_a\mu_a(\mathbf{r})\rho_a(\mathbf{r})+\ln\left[\int\prod_a d\mathbf{r}^{(a)}\exp\left(-\sum_a\mu_a(\mathbf{r}^{(a)})-\beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(K)})\right)\right]\right\}\right). \tag{10}$$

The terms $C_N, C_N', C_N''$ are all constants whose value is unimportant. Now in the limit of large $N$ we use the fact that

$$\int d\mathbf{r}\mu_a(\mathbf{r})\rho_a(\mathbf{r})\to\int d\mathbf{r}\,\mu_a(\mathbf{r})p_a(\mathbf{r})+O\left(\frac{1}{\sqrt{N}}\right), \tag{11}$$

to eliminate the dependence on the quenched disorder in $\rho_a(\mathbf{r})$, and obtain

$$Z_N\approx C_N''\int\prod_a d[\mu_a]\exp\left(N\left\{\int d\mathbf{r}\sum_a\mu_a(\mathbf{r})p_a(\mathbf{r})+\ln\left[\int d\mathbf{r}^{(a)}\exp\left(-\sum_a\mu_a(\mathbf{r}^{(a)})-\beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(K)})\right)\right]\right\}\right). \tag{12}$$

The above integral may be evaluated by the saddle point method. The method works in this limit because the thermodynamics is determined solely by the $p_a$, which are the first moments of the random quenched densities $\rho_a$. We find the following expression for the free energy per number of particles, $N$, in each set $S_a$:

$$-\beta f = \max_{\mu_a}\Bigg\{\int d\mathbf{r}\sum_a \mu_a(\mathbf{r})p_a(\mathbf{r}) + \ln\Bigg[\int d\mathbf{r}^{(a)}\exp\Bigg(-\sum_a \mu_a(\mathbf{r}^{(a)}) - \beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(K)})\Bigg)\Bigg]\Bigg\} + \frac{\gamma_N}{N}. \tag{13}$$

The last term is a $\beta$- and $\mu$-independent term and is related to the infinite-temperature entropy. It does not contribute to the energy of the system, which is what interests us here. The resulting saddle point equation for the $\mu_a$ is

$$p_a(\mathbf{r}_a) = \frac{1}{\mathcal{Z}}\exp[-\mu_a(\mathbf{r}_a)]\int\prod_{b\neq a} d\mathbf{r}^{(b)}\exp\Bigg(-\sum_{b\neq a}\mu_b(\mathbf{r}^{(b)}) - \beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(a)},\ldots,\mathbf{r}^{(K)})\Bigg), \tag{14}$$

where

$$\mathcal{Z} = \int d\mathbf{r}^{(a)}\exp\Bigg(-\sum_a \mu_a(\mathbf{r}^{(a)}) - \beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(K)})\Bigg). \tag{15}$$

One sees in the above that if $\mu_a(\mathbf{r})$ is a solution to the saddle point equation then so is $\mu_a(\mathbf{r})+\alpha_a$, where $\alpha_a$ are arbitrary constants. This is because the constraints $\int d\mathbf{r}\, p_a(\mathbf{r})=1$ are automatically satisfied and so the zero-frequency Fourier modes of the $\mu_a$ are redundant and are zero modes of the theory. We can use one of these free zero modes to set $\mathcal{Z}$ in Eq. (15) to be equal to 1. The saddle point equation then becomes

$$p_a(\mathbf{r}^{(a)}) = \exp[-\mu_a(\mathbf{r}^{(a)})]\int\prod_{b\neq a} d\mathbf{r}^{(b)}\exp\Bigg(-\sum_{b\neq a}\mu_b(\mathbf{r}^{(b)}) - \beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(a)},\ldots,\mathbf{r}^{(K)})\Bigg), \tag{16}$$

and we find that the part of the free energy relevant to the calculation of the energy is

$$\beta f = -\int d\mathbf{r}\sum_a p_a(\mathbf{r})\mu_a(\mathbf{r}), \tag{17}$$

and the average energy is given by $\epsilon=\partial\beta f/\partial\beta$. This can be manipulated to find that, in general, the energy may be written

$$\epsilon = \int\prod_a d\mathbf{r}^{(a)} V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(a)},\ldots,\mathbf{r}^{(K)})\exp\Bigg(-\sum_b \mu_b(\mathbf{r}^{(b)}) - \beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(a)},\ldots,\mathbf{r}^{(K)})\Bigg). \tag{18}$$

We have verified the formalism for nonuniform distributions in some simple cases, but the rest of the paper will be concerned with uniform distributions in domains of unit size. We write $s_a(\mathbf{r})=\exp[\mu_a(\mathbf{r})]$ to obtain the equations in the following form:

$$s_a(\mathbf{r}^{(a)}) = \int\prod_{b\neq a} d\mathbf{r}^{(b)}\exp[-\beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(a)},\ldots,\mathbf{r}^{(K)})]\times\prod_{b\neq a}\frac{1}{s_b(\mathbf{r}^{(b)})},$$

$$\epsilon = \int\prod_a d\mathbf{r}^{(a)} V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(K)})\prod_{b\neq a}\frac{1}{s_b(\mathbf{r}^{(b)})}\times\exp[-\beta V(\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(K)})]. \tag{19}$$

These equations now form the basis for the rest of the paper, which explores their consequences.

In general Eq. (16) cannot be solved analytically; however, in the limit $\beta\to\infty$ it can be simplified by a saddle point method if one writes $s_a(\mathbf{r})=\exp[-\beta w_a(\mathbf{r})]$ which yields

$$w_a(\mathbf{r}) = \min_{\mathbf{r}_1,\ldots,\mathbf{r}_K|\mathbf{r}_a=\mathbf{r}}\Bigg\{V(\mathbf{r}_1,\ldots,\mathbf{r}_a) - \sum_{b\neq a} w_b(\mathbf{r}_b)\Bigg\}. \tag{20}$$

The above-zero-temperature saddle point equation bears a striking similarity to the cavity equations for random link matching [3,4]. This correspondence may eventually be useful in understanding the applicability of random-link-type approximations to site disordered systems. When we are interested in the maximum rather than the minimum problem we can change either the sign of $V$ or $\beta$, changing the sign of $\beta$ simply leads the min above to be changed to a max. In the limit $\beta\to\infty$, using Eq. (17), this yields the ground state energy or optimal matching cost to be

$$\epsilon_{GS} = \sum_a\int d\mathbf{r}\, w_a(\mathbf{r}). \tag{21}$$

The saddle point equations (19) are similar to those studied in our work on Hamiltonian paths but with the major difference that there are now $K+1$ functional order parameters $\mu_a(\mathbf{r})$. In the case where $V$ is a symmetric function and all the $p_a$ are the same for each set, implicitly the case here as we take them all to be uniform, there will clearly be a set symmetric solution where $\mu_a(\mathbf{r})=\mu(\mathbf{r})$. However, it is possible that this symmetry could be spontaneously broken. Thus, despite the fact that the matching problem superficially looks somewhat simpler than the Hamiltonian path problem, it has the potential to exhibit more complex behavior.

The simplest example we can consider is the bipartite matching problem with $K=1$. In this case the saddle point equations read

$$s_0(\mathbf{r}) = \int d\mathbf{r}'\frac{\exp[-\beta V(\mathbf{r},\mathbf{r}')]}{s_1(\mathbf{r}')},$$

$$s_1(\mathbf{r}) = \int d\mathbf{r}'\frac{\exp[-\beta V(\mathbf{r}',\mathbf{r})]}{s_0(\mathbf{r}')}. \tag{22}$$

If one considers the set symmetric solution $s_0(\mathbf{r})=s_1(\mathbf{r})=s_{tsp}(\mathbf{r})$, the resulting equation is

$$s_{tsp}(\mathbf{r}) = \int d\mathbf{r}' \frac{\exp[-\beta V(\mathbf{r},\mathbf{r}')]}{s_{tsp}(\mathbf{r}')}. \tag{23}$$

This is exactly the same equation as that occurring for the TSP problem, but in the TSP problem $s_{tsp}(\mathbf{r})=\exp[\mu(\mathbf{r})/2]$, where $\mu$ is again the Lagrange multiplier fixing the density of points, and the TSP free energy per site is given by

$$\beta f_{tsp} = -\int d\mathbf{r}\,\mu(\mathbf{r}) = -2\int d\mathbf{r}\,\ln[s_{tsp}(\mathbf{r})]. \tag{24}$$

However, we see immediately that for the bipartite matching problem the free energy, within the set symmetric solution, is given by

$$\beta f = -2\int d\mathbf{r}\,\ln[s(\mathbf{r})] = \beta f_{tsp}. \tag{25}$$

Thus the free energy of the TSP with $N$ points is the same as that of a bipartite match between two sets of $N$ points. Notice that in the TSP there are $N$ links and in the bipartite match there are also $N$ links; this equivalence is conjectured to hold for the two minimal versions in the limit where the energy is scaled so as to become extensive as $\beta \to \infty$. If the set symmetric solution is indeed that valid everywhere, our result shows the strict thermodynamic equivalence of the two models at any finite temperature and for an arbitrary distance function.

For pairwise potentials arranged as paths, Eq. (4), and uniform probability density over the domain we may search for general $K$ solutions along the lines of the one found above for the bipartite case. The noncyclic case turns out to be more tractable. We search for a solution in which $s_a(\mathbf{r}) = \exp[\mu_a(\mathbf{r})]$ is the same for all sets except the first ($a=0$) and last ($a=K$) where the ends of the matching path lie:

$$s^{(a)}(\mathbf{r}) = \begin{cases} s(\mathbf{r}), & a=1,\ldots,K-1, \\ s_E(\mathbf{r}), & a=0,K. \end{cases} \tag{26}$$

The saddle point equations now read

$$s_E(\mathbf{r}_0) = \int d\mathbf{r}_1\cdots d\mathbf{r}_K \frac{\exp\left(-\beta\sum_{a=1}^{K} V_P(\mathbf{r}_{a-1},\mathbf{r}_a)\right)}{s(\mathbf{r}_1)\cdots s(\mathbf{r}_{K-1})s_E(\mathbf{r}_K)}, \tag{27}$$

$$s(\mathbf{r}_1) = \int d\mathbf{r}_0 d\mathbf{r}_2\cdots d\mathbf{r}_K \frac{\exp\left(-\beta\sum_{a=1}^{K} V_P(\mathbf{r}_{a-1},\mathbf{r}_a)\right)}{s_E(\mathbf{r}_0)s(\mathbf{r}_2)\cdots s_E(\mathbf{r}_K)}, \tag{28}$$

and the free energy corresponding to this saddle point is

$$\beta f = -\int d\mathbf{r}[2\ln(s_E) + (K-1)\ln(s)]. \tag{29}$$

Using the structure of the pairwise potential, these equations can be solved in terms of a set of $K$ coupled integral equations for the quantities $t_a(\mathbf{r})$:

$$t_0(\mathbf{r}) = \int d\mathbf{r}' \exp[-\beta V_P(\mathbf{r},\mathbf{r}')]\frac{1}{t_{K-1}(\mathbf{r}')} = T\left(\frac{1}{t_{K-1}}\right), \tag{30}$$

$$t_1(\mathbf{r}) = \int d\mathbf{r}' \exp[-\beta V_P(\mathbf{r},\mathbf{r}')]\frac{1}{t_{K-2}(\mathbf{r}')} = T\left(\frac{1}{t_{K-2}}\right), \tag{31}$$

$$t_2(\mathbf{r}) = \int d\mathbf{r}' \exp[-\beta V_P(\mathbf{r},\mathbf{r}')]\frac{t_0(\mathbf{r}')}{t_1(\mathbf{r}')t_{K-2}(\mathbf{r}')} = T\left(\frac{t_0}{t_1 t_{K-2}}\right), \tag{32}$$

$$t_3(\mathbf{r}) = \int d\mathbf{r}' \exp[-\beta V_P(\mathbf{r},\mathbf{r}')]\ \frac{t_0(\mathbf{r}')}{t_2(\mathbf{r}')t_{K-2}(\mathbf{r}')}\Bigg) = T\left(\frac{t_0}{t_2 t_{K-2}}\right), \tag{33}$$

$$\begin{aligned} t_{K-2}(\mathbf{r}) &= \int d\mathbf{r}' \exp[-\beta V_P(\mathbf{r},\mathbf{r}')]\frac{t_0(\mathbf{r}')}{t_{K-3}(\mathbf{r}')t_{K-2}(\mathbf{r}')} \\ &= T\left(\frac{t_0}{t_{K-3}t_{K-2}}\right), \end{aligned} \tag{34}$$

$$t_{K-1}(\mathbf{r}) = \int d\mathbf{r}' \exp[-\beta V_P(\mathbf{r},\mathbf{r}')]\frac{1}{t_0(\mathbf{r}')} = T\left(\frac{1}{t_0}\right), \tag{35}$$

where we have introduced $T$ as the integral operator appearing throughout. Once the $t_a$'s are determined the solution is given by

$$s_E(\mathbf{r}) = t_{K-1}(\mathbf{r}), \tag{36}$$

$$s(\mathbf{r}) = t_0(\mathbf{r})t_{K-2}(\mathbf{r}). \tag{37}$$

This set of equations always has a solution with all $t_a$'s taken the same, $t_a(\mathbf{r}) = s_{tsp}(\mathbf{r})$, this then yields

$$s_E(\mathbf{r}) = s_{tsp}(\mathbf{r}), \tag{38}$$

$$s(\mathbf{r}) = s_{tsp}^2(\mathbf{r}), \tag{39}$$

and inserting this into the free energy we have

$$\beta f = -2K\int d\mathbf{r}\,\ln(s_{tsp}) = K\beta f_{tsp}. \tag{40}$$

The interpretation is clearly that the free energy per link of this matching problem is exactly the same as for the TSP. This result is supported by direct Monte Carlo simulation in the case of a two-dimensional box. This is a generalization of the thermodynamic equivalence seen between the TSP and the bipartite matching problem. Unfortunately, neither for closed paths nor for fully connected matches have we found any general results, and we shall consider only special cases in the following section. Finally we comment that for sufficiently symmetric potentials on closed domains (e.g., with periodic boundary solutions) we find the solution $s_a = \text{const}$ for all $a$. This means that the annealed approximation to the free energy is exact [9] and it is easy to show that $\epsilon_{GS}$ is just

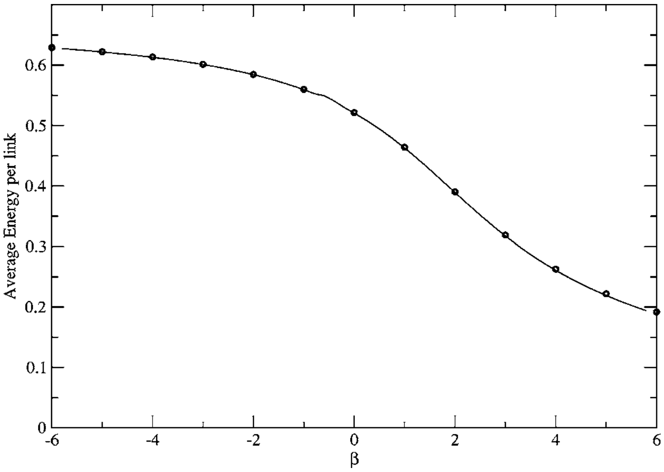

FIG. 5. Tripartite matching with a triangular potential in two dimensions. Points are from Monte Carlo simulations (the error bars are too small to show) and the continuous line is an iterative solution of Eq. (16). Negative temperatures correspond to the maximal problem.

the ground state energy of a single molecule or match with $m=K+1$ sites in the case where each of the $m$ sites is free to move.

## III. TRIPARTITE MATCHING: NUMERICAL RESULTS

Our analytic efforts have not revealed a solution for a symmetric $K=2$ tripartite system other than a constant on symmetric closed domains. We have therefore performed some numerical simulations of the triangle potential. We have considered two domains, a one-dimensional (1D) line and a 2D box, each with boundaries. After preliminary tests to estimate the strength of $1/N$ corrections and the time necessary to equilibrate, we choose the following Monte Carlo parameters: $N=1000$, equilibration for $1\times10^6$ steps, and measurements taken over the succeeding $1\times10^6$ steps. Energies are averaged over 100 samples with different sets of random points. Our codes are checked by showing that the bipartite $K=1$ case agrees with the TSP result.

In Fig. 5 the Monte Carlo results for the energy are confronted with a numerical (iterative) solution of the equations. We find that, in all cases, the iterative approach converges to a symmetric solution with all $s_a(\mathbf{r})$'s the same, even when each $s_a(\mathbf{r})$ is seeded with rather different initial conditions. In two dimensions, the iterative solution is slow for a reasonable discretization of the domain; however, we find that for the temperatures we consider, the solutions are smooth and the procedure gives good results even for a discretization with only 400 points.

We have also considered tripartite matching with the star potential and again see no evidence for breaking of the symmetry between the sets.

## IV. ZERO-TEMPERATURE RESULTS IN ONE AND TWO DIMENSIONS AND CORRESPONDING HEURISTICS

In this section we consider one- and two-dimensional examples where our formalism can solve the original maximal optimization problem. We study Eqs. (20) and (21) and find that, as was the case for the TSP [9], the method of solution throws light on local heuristics that might be used to solve specific instances.

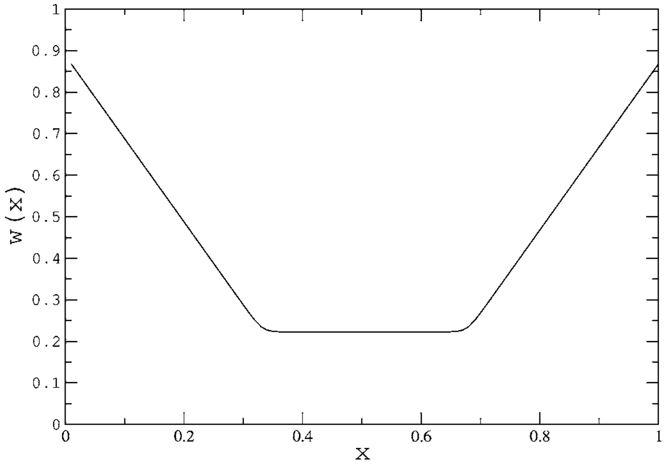

FIG. 6. The function $w(x)$ for tripartite matching on a one-dimensional line, obtained by numerically solving Eq. (19) for $\beta=70.0$. The parameters agree with the solution based on a greedy heuristic discussed in the text.

First consider maximal tripartite matching in one dimension with a triangular potential based on a pairwise potential that is simply the distance between the points,

$$V_{triangle}(x_0,x_1,x_2) = |x_0-x_1| + |x_1-x_2| + |x_2-x_0| = 2[\max(x_0,x_1,x_2) - \min(x_0,x_1,x_2)]. \tag{41}$$

Evidently the location of the middle point does not contribute to the potential. We search for a piecewise linear function $w(x)$ consisting of three linear sections evenly dividing the unit interval as shown in Fig. 6. A $1/3$ jump heuristic guesses that the maximum occurs for points arranged as shown in Fig. 7. This basically means that the match consists of a central point in the central third of the interval, matched with points in the first and last third, each at a distance of a third from the central point. In this case the Eq. (20) can be written

$$w(x) + w(x+1/3) + w(x+2/3) = V_{triangle}(x,x+1/3,x+2/3) = 4/3, \tag{42}$$

where $0<x<1/3$. This of course immediately gives the average value of $4/3$ per match in the optimal match. We now make a piecewise linear anzatz for $w(x)$ (which is symmetric about $x=1/2$)

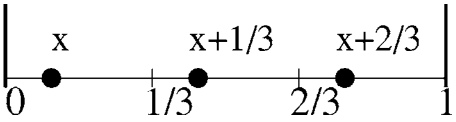

FIG. 7. Arrangement of points for the one-third jump heuristic.

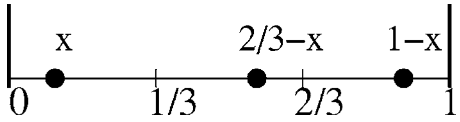

FIG. 8. Arrangement of points for the greedy heuristic.

$$w(x) = \begin{cases} a\left|x - \dfrac{1}{2}\right| + b & \text{for } x \in \left[0, \dfrac{1}{3}\right] \cup \left[\dfrac{2}{3}, 1\right], \\ a'\left|x - \dfrac{1}{2}\right| + b' & \text{for } x \in \left[\dfrac{1}{3}, \dfrac{2}{3}\right]. \end{cases} \tag{43}$$

Immediately one sees from Eq. (42) that we must have $a'=0$. Now substituting this into Eq. (20) we find

$$\frac{4}{3}a + 2b + b' = \frac{4}{3}. \tag{44}$$

The coefficient $a$ can be determined by looking locally at Eq. (20) when $x \in [\frac{1}{3}, \frac{2}{3}]$ and, writing $x_1 = x - \frac{1}{3} + \epsilon_1$ and $x_2 = x + \frac{1}{3} + \epsilon_2$, this gives

$$b' = \max_{\epsilon_1, \epsilon_2}\left\{\frac{4}{3} + 2(\epsilon_2 - \epsilon_1) - a(\epsilon_2 - \epsilon_1) - \frac{2}{3}a - 2b\right\} \tag{45}$$

and the local stationarity with respect to $\epsilon_1$ and $\epsilon_2$ then gives us that $a=2$. Now we assume continuity of $w$, which gives

$$b' = \frac{2}{9}. \tag{46}$$

This thus yields the solution for Eq. (20) in complete agreement with the numerical resolution of the low-temperature saddle point equations shown in Fig. 6.

An alternative heuristic for this triangular distance potential is the greedy heuristic with the arrangement of points shown in Fig. 8. In this case the maximization equation is

$$\begin{aligned} w(x) + w(2/3 - x) + w(1 - x) &= V_{triangle}(x, 2/3 - x, 1 - x) \\ &= 2(1 - 2x). \end{aligned} \tag{47}$$

Interestingly, despite the fact that the heuristic here is different it yields the same function $w(x)$, and consequently the same average value per match of the optimal match. This is presumably a peculiarity of linear pairwise potentials in one dimension, and a similar phenomenon is seen for the maximal TSP [9].

These issues generalize to other cyclic (and possibly fully connected) potentials in one-dimensional $m$-partite matching. The star potential based on the same distance metric does not lead to such a simple solution as the locations of all three points are important and $w(x)$ is no longer piecewise linear.

Now, remaining in one-dimension, consider the triangle potential built from a quadratic pairwise potential. The locations of all three points are needed to determine the cost of a match (in contrast to the linear distance function), and it is harder to guess a heuristic. However, numerical experiments hint that the solution $w(x)$ is quadratic so we try the anzatz $w(x) = a + b(x - \frac{1}{2})^2$ (which respects the symmetry about the midpoint $x = 1/2$) in the maximization equation

$$w(x) = \max_{x_1, x_2}[(x_1 - x)^2 + (x_2 - x)^2 + (x_1 - x_2)^2 - w(x_1) - w(x_2)]. \tag{48}$$

The maximization equations for the right hand side yield

$$\begin{aligned} (2 - b)x_1 - x_2 &= x - \frac{1}{2}b, \\ (2 - b)x_2 - x_1 &= x - \frac{1}{2}b. \end{aligned} \tag{49}$$

A manifestly nonoptimal solution to the above is $x_2 = x_3$ and to avoid this we require the above equation to have more than one solution; this means that $b=1$ or 3. The choice $b=1$ yields $x=1/2$ which is clearly generally not the case. The choice $b=3$ yields $x + x_1 + x_2 = 3/2$ which can be written as

$$\frac{1}{3}(x + x_1 + x_2) = \frac{1}{2}, \tag{50}$$

meaning that the center of mass of optimal triangles is in the center of the interval. Using this condition in the maximal equation, all parameters are determined, we find that $a=0$ and the energy is given by $3/4$ which can further be checked by numerical iteration of Eq. (16). The condition above is not a full heuristic in that it does not determine the location of all points in the match when given just one. However, it is supported by Monte Carlo simulations where the sum of the coordinates of all matching triangles is a bell-shaped distribution with correct mean, and variance decreasing rapidly with temperature. We expect this result to generalize to $m$-partite matching with a cyclic potential with quadratic pairwise $V_P$. It also generalizes to a star potential with quadratic $V_P$ where we find energy $1/4$. Moreover, although it is not an exact solution, the quadratic form of $w(x)$ is a good approximation even at quite elevated temperatures.

The amenability of quadratic potentials carries over to two dimensions. We consider tripartite matching with a triangle potential in a disk so as to preserve rotational symmetry and take $w(\mathbf{r})$ to be independent of angle. The function to be maximized is smooth, and by differentiating we find that a symmetric maximum occurs when the center-of-mass condition $\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3 = \mathbf{0}$ holds. As in the one-dimensional case this is not a full heuristic; however, it is sufficient to determine the energy, which is $9/2\pi$. This is consistent with numerical iteration of Eq. (16) at low temperature, though in two dimensions numerical resolution of the saddle point at low temperatures is difficult. Furthermore Monte Carlo simulations indicate that the center -of-mass condition indeed becomes sharper as the temperature is reduced. A specific heuristic which yields this energy makes greedy matches based on equilateral triangles with center at the origin. The same heuristic can be used in the case of a triangle potential in two dimensions based on a linear distance pairwise potential; it leads to an energy of $2\sqrt{3/\pi}$, but although this is in good agreement with numerical results we have not been able to prove that it is the true maximum.

## V. CONCLUSIONS

We have presented a statistical-mechanics-based approach to multi-index matching problems with site disorder, specifically when the points to be matched are randomly distributed in a domain $\mathcal{D}$. The cost functions considered are functions of the relative distances between the points and are consequently correlated, in contrast to the random link model. We have analyzed the $m$-partite version of these problems where each match contains one element from $m$ distinct sets, each of which can in principle have different probability distributions in the domain $\mathcal{D}$. In the thermodynamic limit the system is described by $m$ functional order parameters. We have concentrated on maximal matching problems as the ground state energy in this case is extensive with the scaling we employ.

In the special cases where the cost function is sufficiently symmetric and the probability distributions of each set are the same we find set symmetric solutions where the functional order parameter of each set is the same. We have not found any evidence for a breaking of this symmetry; an open question is whether there are geometries (e.g., effects of boundaries and spatial dimensionality) where this symmetry is spontaneously broken.

For open chain potentials (which of course include bipartite matching) we find a solution to the saddle point equations that gives exactly the same free energy as the traveling salesman problem with the same number of links. Furthermore, this equivalence holds at all temperatures and is independent of the precise functional form of the pairwise potential from which the chain potential is constructed. Interestingly this observation is analogous to an equivalence that is conjectured to hold between Euclidean bipartite matching and the Euclidean TSP in the zero-temperature limit of the minimal problem, where the cost function is scaled to give an extensive ground state energy.

As in the TSP we find that the zero-temperature saddle point equations can be solved via *Ansätze* inspired by heuristics to find the optimal match. We have been able to analytically solve the saddle point equations for the function $w$ in a number of instances of maximal matching in one and two dimensions. Also we have numerically verified that the average optimal match is accurately predicted and that the heuristic that was used to solve the saddle point equation is indeed that associated with the optimal match. Finally, we remark that knowledge of the function $w$ does not completely specify a heuristic; however, it does give some partial, potentially useful, information about the heuristic. Indeed, we have demonstrated that two completely different heuristics can give the same average of the optimal match for a simple Euclidean tripartite-matching problem in one dimension.

## ACKNOWLEDGMENTS

D.S.D. would like to thank the Isaac Newton Insitute, Cambridge University, where part of this work was carried out during the program Principles of the Dynamics of Non-Equilibrium Systems. D.L. acknowledges the EPSRC.

[1] C. H. Papadimittiou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Dover, New York, 1999).

[2] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[3] M. Mézard and G. Parisi, J. Phys. (France) Lett. **46**, L771 (1985); Europhys. Lett. **2**, 913 (1986); J. Phys. (France) **49**, 2019 (1988).

[4] O. C. Martin, M. Mézard and O. Rivoire, Phys. Rev. Lett. **93**, 217205 (2004); J. Stat. Mech.: Theory Exp. (2005) P09006.

[5] A. B. Poore, Comput. Optim. Appl. **3**, 27 (1994).

[6] M. Johnsson, G Magyar, and O. Nevalaien, Nord. J. Comput. **5**, 143 (1998); G. Magyar, M. Johnsson, and O. Nevalainen, Acta Cybern. **14**, 357 (1999).

[7] M. Weber and Th. M. Liebling, Zeitschrift fur Operations Research **30**, 85 (1986).

[8] J. Houdayer, J. H. Boutet de Monvel, and O. C. Martin, Eur. Phys. J. B **6**, 383 (1998); J. H. Boutet de Monvel and O. C. Martin, Phys. Rev. Lett. **79**, 167 (1997).

[9] D. S. Dean, D. Lancaster, and S. N. Majumdar, J. Stat. Mech.: Theory Exp. (2005) L01001; Phys. Rev. E **72**, 026125 (2005).