COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA 22 (4) (2002) 523–530

ALMOST SURE CONVERGENCE OF THE MINIMUM BIPARTITE MATCHING FUNCTIONAL IN EUCLIDEAN SPACE

J. H. BOUTET DE MONVEL, O. C. MARTIN

Received 25 February, 1997

Let $L_N = L_{MBM}(X_1,\ldots,X_N;Y_1,\ldots,Y_N)$ be the minimum length of a bipartite matching between two sets of points in $\mathbf{R}^d$, where $X_1,\ldots,X_N,\ldots$ and $Y_1,\ldots,Y_N,\ldots$ are random points independently and uniformly distributed in $[0,1]^d$. We prove that for $d\geq 3$, $L_N/N^{1-1/d}$ converges with probability one to a constant $\beta_{MBM}(d)>0$ as $N\to\infty$.

1. Introduction and statement of the result

Given two sets of N points $X=\{X_1,\ldots,X_N\}$ and $Y=\{Y_1,\ldots,Y_N\}$ in $\mathbf{R}^d$, a bipartite matching of X and Y is a perfect matching M on the set $X\cup Y$, such that each pair in M is made of one point of X and one point of Y. The length of such a matching is defined to be the sum of the euclidean lengths of the edges formed by its pairs. The (euclidean) minimum bipartite matching problem (MBMP) then asks one to find a bipartite matching of X and Y whose length is as small as possible. We shall denote by $L_{MBM}(X,Y)$ the length of a minimum bipartite matching of X and Y.

A related problem is the simple minimum matching problem (MMP), where one is asked to find a perfect matching of smallest euclidean length on a set $X=\{X_1,\ldots,X_N\}\subset\mathbf{R}^d$. The subadditive methods inaugurated by Beardwood, Halton and Hammersley (BHH) [4] and further developed in [12,10,9], show that a strong limit theorem applies to the length $L_{MM}(X)$ of a simple minimum matching on X, when the points $X_1,\ldots,X_N$ are random. The theorem states that for any dimension d, if $X_1,\ldots,X_N,\ldots$ is a

Mathematics Subject Classification (2000): 60D05

0209–9683/102/$6.00

sequence of points distributed independently and uniformly in a bounded region $\Omega \subset \mathbf{R}^d$, then the ratio $L_{MM}(X_1,\dots X_N)/N^{1-1/d}$ converges almost surely to $\mathrm{Vol}(\Omega)^{1/d}\beta_{MM}(d)$, where $\mathrm{Vol}(\Omega)$ denotes the Lebesgues measure of Ω and $\beta_{MM}(d)>0$ is a universal constant depending only upon d.

The functional L_{MBM} does not satisfy this form of limit theorem in dimensions 1 and 2. For $d=1$, the MBMP amounts to a sorting problem and it is not difficult to show that if X and Y both consist of N points independently and uniformly distributed in $[0,1]$, there are constants $0<C_1<C_2$ such that $C_1\sqrt{N} \le L_{MBM}(X,Y) \le C_2\sqrt{N}$ with probability $1-o(1)$ as $N\to\infty$. Moreover in that case the variance of $L_{MBM}(X,Y)/\sqrt{N}$ does *not* converge to zero as $N\to\infty$. (L_{MBM} is not "self-averaging", in the statistical physics' terminology.) For $d=2$ Ajtai et al. [1] proved a remarkable fact: if the sets X,Y are now distributed in $[0,1]^2$, then for some constants C_1,C_2 indendent of N, one has $C_1\sqrt{N\log N} \le L_{MBM}(X,Y) \le C_2\sqrt{N\log N}$ with probability $1-o(1)$. Numerical simulations suggest that $L_{MBM}(X,Y)/\sqrt{N\log N}$ converges to a non-random constant as $N\to\infty$, however this has not yet been proved.

In this article, we show that for any $d\ge 3$ we recover a BHH theorem for the functional L_{MBM}.

Theorem 1.1. *Let $X_1,\dots,X_N,\dots$ and $Y_1,\dots,Y_N,\dots$ be two sequences of random points independently and uniformly distributed in $[0,1]^d$, where $d\ge 3$, and let $L_N=L_{MBM}(X_1,\dots,X_N;Y_1,\dots,Y_N)$. There exists a constant $\beta_{MBM}(d)>0$ such that with probability one*

$$\lim_{N\to\infty} L_N/N^{1-1/d} = \beta_{MBM}(d).$$

2. Proof of Theorem 1.1

To begin, we remark that to prove this theorem it will suffice to establish that $L_N/N^{1-1/d}$ converges in mean value to a constant $\beta_{MBM}(d)$. This is a consequence of the following lemma [14]:

Lemma 2.1. *For any $t>0$, one has*

$$P\left(\left|\frac{L_N}{N^{1-1/d}} - E\Big(\frac{L_N}{N^{1-1/d}}\Big)\right| > t\right) \le 2\exp\Big(-\frac{N^{1-2/d}t^2}{8d}\Big).$$

This result follows from the application of Azuma's inequality [3] and the martingale difference method to L_N, in a way by now standard in the probabilistic theory of combinatorial optimisation [13]. Given the lemma, the theorem follows easily from the convergence of $EL_N/N^{1-1/d}$ as $N\to\infty$, by applying the Borel-Cantelli lemma.

We have now to establish that for $d\geq 3$ the quantity $EL_N/N^{1-1/d}$ indeed converges to a constant $\beta_{MBM}(d)>0$. To prove this we exploit the subadditivity properties of L_{MBM}, in the spirit of Steele's theory of subadditive Euclidean functionals [12]. Let us divide the unit cube $[0,1]^d$ into disjoint similar subcubes Q_k, $k=1,\dots,m^d$ with edges of length $1/m$, and compare the value of $L_{MBM}(X,Y)$ to the sum

$$\sum_{k=1}^{m^d} L_k, \tag{1}$$

where L_k is the value of the functional L_{MBM} for the set of points X_i and Y_i which belongs to Q_k. A difficulty arises as in general the Q_k's do not contain the same number of points X_i and of points Y_i. (In fact the special properties of the MBMP in dimensions 1 and 2 originate from the fluctuations of the differences between these numbers around their mean value 0.) To give meaning to the sum (1) we need to generalize the functional L_{MBM} to matchings between two sets of different cardinalities. There are several ways to do this; we shall define $L_{MBM}(X_1,\dots X_{N_1};Y_1,\dots Y_{N_2})$ by imposing that the minimum matching contains as few unmatched points as possible. That is if $N_1>N_2$, we leave N_1-N_2 points of X unmatched, whereas if $N_1<N_2$ we leave N_2-N_1 points of Y unmatched.

Although expression (1) now makes sense, it is still not possible to write a subadditivity inequality of the same form as the one studied in [12]. Indeed, such a form (which Steele calls "geometric subadditivity") implies an upper bound of the form $CN^{1-1/d}$ for the functional at hand [13], and it is easy to see that no such bound applies to $L_{MBM}(X,Y)$. We shall however see that a geometric subadditivity property holds *in the mean* for the functional L_{MBM}. Suppose that the points $X_1,\dots X_{N_1},Y_1,\dots Y_{N_2}$ belong to an arbitrary cube Q having edge length a, and divide Q into disjoint cubes Q_p, $p=1,\dots 2^d$ by splitting each edge in two halves. Construct in each Q_p an optimal matching in the sense just defined, between the $n_{1,p}$ points X_i and the $n_{2,p}$ points Y_i in Q_p, and denote its length by L_p. The points that are left unpaired are in number $|n_{1,p}-n_{2,p}|$ in each Q_p, so if L_0 denotes the length of an optimal matching for these points one has

$$
\begin{aligned}
L_{MBM}(X_1,\dots X_{N_1};Y_1,\dots,Y_{N_2}) &\le \sum_{p=1}^{2^d} L_p + L_0 \\
&\le \sum_{p=1}^{2^d} L_p + \frac{1}{2}a\sqrt{d}\sum_{p=1}^{2^d}|n_{1,p}-n_{2,p}|, \qquad (2)
\end{aligned}
$$

where the last inequality is obtained by bounding L_0 in an obvious way.

We shall apply this to $Q=[0,1]^d$. Let Q_{p_1} $p_1=1,\dots 2^d$ be the cubes obtained in the above subdivision; let $Q_{p_1p_2}$ be the cubes obtained by splitting in two halves the edges of each cube Q_{p_1}; and so on. By repeating this operation K times, we get a subdivision with cubes $Q_{p_1\dots p_K}$ whose edges are of length $1/2^K$. Let $n_{1,p_1\dots p_K}$ and $n_{2,p_1\dots p_K}$ be respectively the number of points X_i and Y_i in $Q_{p_1\dots p_K}$. Apply (2) first to the $Q_{p_1,\dots p_{K-1}}$'s, then to the $Q_{p_1\dots p_{K-2}}$'s, etc, keeping at each step only those points which are still unpaired. It is easy to convince oneself that the number of unpaired points in each $Q_{p_1,\dots p_{K-k}}$ just after step k is given by $|n_{1,p_1,\dots p_{K-k}}-n_{2,p_1,\dots p_{K-k}}|$. After step $k=K$ one obtains a matching between $X_1,\dots X_{N_1}$ and $Y_1,\dots Y_{N_2}$ where all the points but $|N_1-N_2|$ are matched. One is thus led to the following inequality:

$$
\begin{aligned}
L_{MBM}(X_1,\dots X_{N_1};Y_1,\dots Y_{N_2}) &\le \sum_{p_1\dots p_K} L_{p_1\dots p_K} \\
&+\sum_{k=1}^{K}\frac{\sqrt{d}}{2^k}\sum_{p_1\dots p_k}|n_{1,p_1\dots p_k}-n_{2,p_1\dots p_k}|. \qquad (3)
\end{aligned}
$$

We now proceed to derive a subadditivity property for the mean value of $L_{MBM}(X,Y)$. We first consider the case where $N_1=\text{cardX}$ and $N_2=\text{cardY}$ are not fixed integers but are independent Poisson random variables with the same mean value N, the elements of X and Y being chosen independently and uniformly in $[0,1]^d$. For a given k, the numbers $n_{1,p_1,\dots p_k}$ and $n_{2,p_1,\dots p_k}$ are then also independent Poisson random variables, with parameter $N/2^{kd}$. Let $M(N)=EL_{MBM}(X_1,\dots X_{N_1};Y_1,\dots Y_{N_2})$. It is immediate by homogeneity that

$$
EL_{p_1\dots p_K}=2^{-K}M(N/2^{Kd}). \qquad (4)
$$

Moreover from the well known properties of Poisson variables we have

$$
E|n_{1,p_1\dots p_k}-n_{2,p_1\dots p_k}|\le\sqrt{2}\Big(\frac{N}{2^{kd}}\Big)^{1/2}. \qquad (5)
$$

By taking mean values in (3) we obtain:

$$M(N) \leq 2^{K(d-1)} M(N/2^{Kd}) + \sqrt{2dN} \sum_{k=1}^{K} 2^{k(d/2-1)}. \tag{6}$$

This inequality has been obtained for a subdivision of $[0,1]^d$ which consists in 2^{Kd} similar cubes. Suppose now that we start from the subdivision Σ in m^d similar cubes Q_k $k=1,\dots m^d$, where m is an arbitrary integer. One can then reproduce the previous construction in the following manner. Let $m=2^K+r$ where $0 \leq r < 2^K$. Consider the cube $Q_0=[0,2^{K+1}/m]^d$ and form the natural subdivision Σ_0 of Q_0 by $2^{(K+1)d}$ cubes $Q_{p_0,\dots p_K}$ whose edges have length $1/m$. We can proceed with Q_0 and Σ_0 to a $K+1$ steps construction similar to the one which led to (3). The only differences are that Q_0 has edges of length $2^{K+1}/m$ rather than 1, and that some of the $Q_{p_0\dots p_K}$'s, namely those which belong to Σ_0 but not to Σ, are empty. Nevertheless, we may write

$$\begin{aligned} & L_{MBM}(X_1,\dots X_{N_1};Y_1,\dots,Y_{N_2}) - \sum_{p=1}^{m^d} L_k \\ & \leq \sum_{k=0}^{K} \frac{\sqrt{d}2^{K-k}}{m} \sum_{p_0\dots p_k} |n_{1,p_0\dots p_k} - n_{2,p_0\dots p_k}| \\ & \leq \sum_{k=0}^{K} \frac{\sqrt{d}}{2^k} \sum_{p_0\dots p_k} |n_{1,p_0\dots p_k} - n_{2,p_0\dots p_k}|. \end{aligned} \tag{7}$$

Now $n_{1,p_0\dots p_k}$ and $n_{2,p_0\dots p_k}$ are Poisson variables with parameter lower than $2^{(K-k)d}N/m^d \leq 2^{-kd}N$ so we still have

$$E|n_{1,p_0\dots p_k} - n_{2,p_0\dots p_k}| \leq \sqrt{2}\Big(\frac{N}{2^{kd}}\Big)^{1/2}. \tag{8}$$

Taking average values one is led to

$$M(N) \leq m^{d-1} M(N/m^d) + 2^d\sqrt{2dN} \sum_{k=0}^{K} 2^{k(d/2-1)}. \tag{9}$$

Dividing this last inequality by $N^{1-1/d}$ and then replacing N by $m^d N$, we get

$$\frac{M(m^dN)}{(m^dN)^{1-1/d}} \leq \frac{M(N)}{N^{1-1/d}} + \frac{2^d\sqrt{2d}}{N^{1/2-1/d}} \sum_{k=0}^{K} 2^{-k(d/2-1)}. \tag{10}$$

If $d>2$, the sum on the r.h.s. of the last inequality is bounded above independently of N, and is divided by a positive power of N. Elementary analysis now shows that the ratio $M(N)/N^{1-1/d}$ necessarily converges to a limit $\beta_{MBM}(d)$ as $N \to \infty$. Indeed, let $f(t) = M(t^d)/t^{d-1}$. One verifies at once that $f(t)$ satisfies

$$f(mt) \leq f(t) + C/t^{d/2-1} \tag{11}$$

for all $t>0$ and any integer m; $f(t)$ is continuous, since $M(N)$ is a continuous function of N. So the expression $f(t)+C_d/t^{d/2-1}$ is bounded in $[1,2]$ and since $[1,\infty[$ is the union of the intervals $m[1,2], m \geq 1$, it follows from (11) that $f(t)$ remains bounded as $t\to\infty$, thus $\lim^* f(t)<\infty$. Now define $\beta=\lim_* f(t)$. For any $\epsilon>0$, chose $t_0 \gg 1$ and $\eta>0$ such that $f(t)+C_d/t^{d/2-1}<\beta+\epsilon$ for t in the interval $I=[t_0-\eta, t_0+\eta]$. Since the intervals mI, $m\geq 1$ span a whole interval $[A,\infty[$ for an A sufficiently large, it follows again from (11) that $\lim^* f(t) \leq \beta+\epsilon$. Since ϵ is arbitrary one has $\lim^* f(t)=\beta$, hence $f(t)\to\beta$ as $t\to\infty$, from which it follows that $\lim_{N\to\infty} M(N)/N^{1-1/d}=\beta$. ∎

We have thus shown for $d\geq 3$, that one has

$$EL_{MBM}(X_1,\ldots,X_{N_1};Y_1,\ldots,Y_{N_2}) \sim \beta^E_{MBMP}(d)N^{1-1/d}, \; N \to \infty \tag{12}$$

when N_1 and N_2 are independent Poisson variables with parameter N. The same result for the mean value EL_N, where N is a fixed integer, follows then easily. Indeed, we have the obvious bound

$$\begin{aligned} &|L_{MBM}(X_1,\ldots X_N;Y_1,\ldots Y_N) - L_{MBM}(X_1,\ldots X_{N_1};Y_1,\ldots Y_{N_2})| \\ &\qquad \leq \sqrt{d}(|N_1-N|+|N_2-N|), \end{aligned} \tag{13}$$

whence taking mean values,

$$|EL_N - EL_{MBM}(X_1,\ldots X_{N_1};Y_1,\ldots Y_{N_2})| \leq 2\sqrt{2dN}, \tag{14}$$

and dividing by $N^{1-1/d}$ we deduce that

$$\lim_{N\to\infty} \frac{EL_N}{N^{1-1/d}} \to \beta_{MBM}(d). \tag{15}$$

Theorem 1.1 is now proved.

3. Concluding remarks

1) Our decimation procedure does not give back the bounds proven by Ajtai *et al.* in $d=2$, but a weaker $O(\sqrt{N}\ln N)$ bound. It is believed that a self-averaging theorem applies also to the functional L_{MBM} in dimension 2 [11].
2) The estimation of the constants $\beta_{MBM}(d)$ is also an interesting problem. A remarkable result of Talagrand [14] shows that one has $\beta_{MBM}(d)=$

$\sqrt{d/2e\pi}(1+O(\ln d/d))$ as $d\to\infty$. It is conjectured that a $1/d$ series expansion actually exists for $\beta_{MBM}(d)$.
3) Mézard and Parisi have obtained detailed analytic predictions for the *random link* versions of the MMP and the MBMP [8], where the distance matrix between the points X_i and Y_j is replaced by a matrix of independent and identically distributed entries. (Some of these predictions, for the random assignment problem, have been proven recently by Aldous [2].) Numerical studies [6,7] indicate that for the MMP and the MBMP, the random link model provides one with a very good "mean-field" approximation to the Euclidean model in the large d limit. Except for simpler combinatorial problems however [5], very few rigorous results are known for comparing the euclidean and the random link models.

Aknowledgments. It is a pleasure to thank J.M. Steele for fruitful discussions and pointing to us reference [14].

References

[1] M. Ajtai, J. Komlós, and G. Tusnády: On optimal matchings, *Combinatorica*, **4** (1984), pp. 259–264.

[2] D. Aldous: The $\zeta(2)$ limit in the random assignment problem, *Random Structures and Algorithms*, **18** (2001), pp. 381–418.

[3] X. Azuma: Weighted sums of certain dependent random variables, *Tohuku Math. Journal*, **19** (1967), pp. 357–367.

[4] J. Beardwood, J. H. Halton, and J. M. Hammersley: The Shortest Path through Many Points, *Proceedings of the Cambridge Philosophical Society*, **55** (1959), pp. 299–327.

[5] D. J. Bertsimas and G. van Ryzin: An Asymptotic Determination of the Minimum Spanning Tree and Minimum Matching Constants in Geometrical Probability, *Operation Research Letters*, **9** (1990), pp. 223–231.

[6] J. H. Boutet de Monvel and O. C. Martin: Mean Field and Corrections for the Euclidean Minimum Matching Problem, *Physical Review Letters*, **79-1** (1997), pp. 167–170.

[7] J. Houdayer, J. Boutet de Monvel, and O. Martin: Comparing Mean Field and Euclidean Matching Problems, *Eur. Phys. Jour. B, cond-mat/9803195*, **6-3** (1998), pp. 383–393.

[8] M. Mézard and G. Parisi: On the Solution of the Random Link Matching Problems, *Journal de Physique*, **48** (1987), pp. 1451–1459.

[9] C. Redmond and J. Yukich: Limit Theorems and Rates of Convergence for Euclidean Functionals, *The Annals of Applied Probability*, **4-4** (1994), pp. 1057–1073.

[10] W. T. Rhee: A Matching Theorem and Subadditive Euclidean Functionals, *Annals of Applied Probability*, **3-3** (1993), pp. 794–801.

[11] W. D. Smith: Studies in Computational Geometry Motivated by Mesh Generation, PhD Thesis, Princeton University, Princeton, NJ, 1989.

[12] M. Steele: Subadditive Euclidean Functionals and Nonlinear Growth in Geometric Probability, *The Annals of Probability*, **9-3** (1981), pp. 365–376.

[13] M. Steele: Probability Theory and Combinatorial Optimisation, SIAM Philadelphia, ISBN 0-89871-380-3, 1997.

[14] M. Talagrand: Matching random samples in many dimensions, *Annals of Applied Probability*, **2** (1992), pp. 846–856.

J. H. Boutet de Monvel

Center for Hearing
and Communication Research,
Karolinska Institutet,
17176 Stockholm, Sweden;
j.boutet.de.monvel@cfh.ki.se

O. C. Martin

Laboratoire de Physique Théorique
et Modèles Statistiques,
Université de Paris-Sud,
91405 Orsay, France;
martino@ipno.in2p3.fr