Notes on Random Euclidean Bipartite Matching Problems

Dario Trevisan
Università degli studi di Pisa,
dario.trevisan@unipi.it

September 8, 2021

Abstract

These are lecture notes for the Master Classes en probabilités at IMRA, Université de Strasbourg, January 20-24th 2020. They provide a rather gentle introduction to the problem of Random Bipartite Matching, aimed at master students (M1) in mathematics. We thank the organizers and in particular N. Juillet for many suggestions while translating the content in French. We also thank all the students for finding many typos and occasional errors.

1 Introduction

Let $d \geq 1$, and consider $2n \geq 1$ i.i.d. random variables

$$(X_i)_{i=1}^n, (Y_i)_{i=1}^n$$

taking values in the cube $[0,1]^d$ with *uniform distribution*, i.e.,

$$P(X_i \in A) = |A|, \quad \text{for every } A \subseteq [0,1]^d \text{ Borel,}$$

and $|A|$ denotes (the d-dimensional) Lebesgue measure of A. We can think of $(X_i)_i$ as red points and $(Y_i)_i$ as blue points (Figure 1). The *Random Euclidean Bipartite Matching Problem* is defined as the following random variational problem,

$$\min_{\sigma \in \mathcal{S}^n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|,$$

where $\mathcal{S}^n$ denotes the set of permutations over $\{1, \ldots, n\}$, and $|\cdot|$ denotes the Euclidean norm:

$$|x| = \sqrt{\sum_{k=1}^d x_k^2}, \quad \text{for } x = (x_k)_{i=1}^d \in \mathbb{R}^d.$$

In other words, we look for a correspondence (or "matching") between red and blue points (given by a permutation σ) in such a way that the sum of the distances between connected points is minimized.

Aim of this lectures is to give an introduction to some features of this problem, in particular when n becomes large. Actually, we slightly generalize allowing for a power of the distance: given $p \in (0, +\infty)$, we define the random variable

$$B_{p,n} := \min_{\sigma \in \mathcal{S}^n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^p. \tag{1.1}$$

We will address natural questions such as the asymptotic behaviour of the random variables $B_{p,n}$ and the structure of matching associated to (random) minimizers $\sigma_{p,n}$,

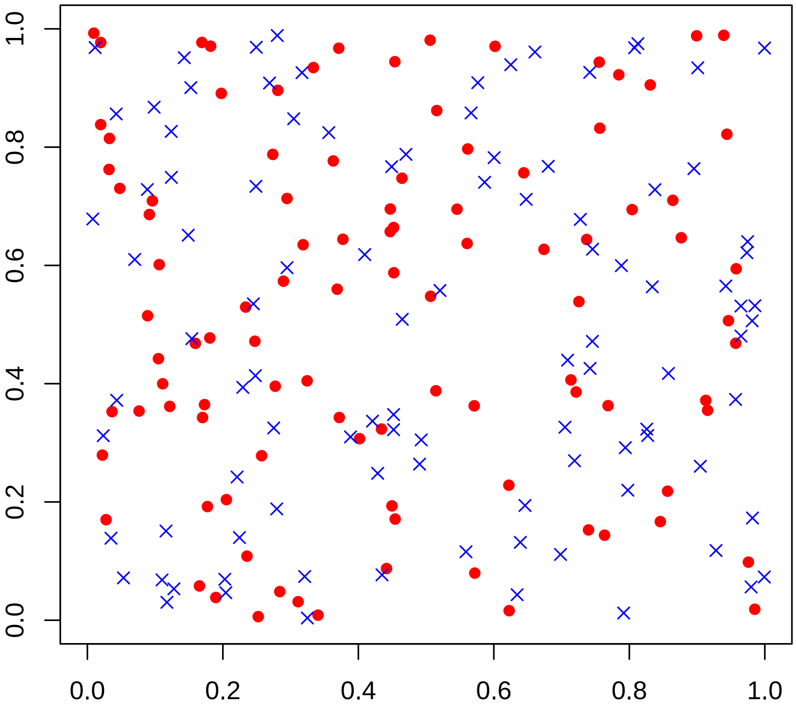


Figure 1: a sample of $n = 100$ pairs of i.i.d. uniform points on a unit square.

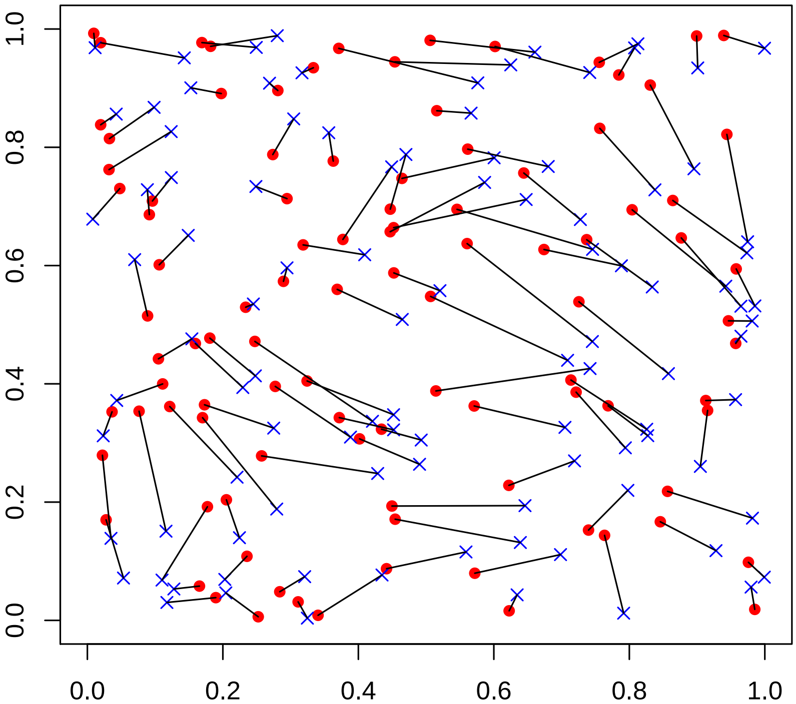


Figure 2: the associated optimal matching with $p = 2$ (below).

i.e., the maps $(x_i, y_{\sigma_{p,n}(i)})_{i=1}^n$ (Figure 2). We will see however that many problems are in fact open already at the level of asymptotics for $B_{p,n}$.

This and related *random combinatorial optimization* problems have been investigated by many authors, as they naturally appear in modelling several phenomena. Moreover, as a general philosophy, one may expect that a "generic" (hence random) instance of a problem exhibits some less peculiar features with respect to specific ones (usually the "worst case scenarios") and one may devise/analyse suitable algorithms.

We can consider many variants of the bipartite matching problem, not only by allowing for more general domains than the cube, but also introducing correlations between the random variables. We will also consider a variant where the numbers of red and blue points are different. In this case however one has to be precise on what a matching is. Depending on the interpretation, this may vary: e.g., [STAD13] we could think that red points are mobile phone users and blue points are antennas (typically much less than the number of users) , so that it is natural then to impose some constraint on the number of users that can connect to each antenna.

Another closely related random combinatorial optimization problem is the *minimal matching* problem, where one considers $2n$ i.i.d. variables $(X_i)_{i=1}^{2n}$ and looks for

$$M_{p,n} = \min_{\sigma \in \mathcal{S}^{2n}} \sum_{i=1}^{n} |X_{\sigma(i)} - X_{\sigma(n+i)}|^p,$$

i.e., one does not prescribe a color of the $2n$ points and simply tries to connect pairs of them. We may call this problem *monopartite* matching when comparing it with the bipartite one (1.1). For an introductory monograph on random combinatorial optimization problems, we suggest [Ste97].

Exercise 1.1. Show that $\mathbb{E}[M_{p,n}] \leq \mathbb{E}[B_{p,n}]$.

2 Lower bounds

Let us argue heuristically first. Since red/blue points are "uniformly distributed" on $[0,1]^d$, the typical configuration should be such that for each X_i one can find Y_j within a distance approximately $1/n^{1/d}$ (thinking e.g. of a grid of evenly spaced points).

This leads to the conjecture

$$B_{p,n} \sim n \cdot \frac{1}{n^{p/d}} = n^{1-p/d}.$$

However, looking at simulations (Figure 3) it appears that at least in the case $d = 1$, $p = 1$, the guess may not be correct, for many distances are larger. A striking feature of the bipartite problem is that indeed the guess is correct (fixing $p = 1$) only if $d \geq 3$, as first showed in [AKT84], although the case $d = 2$ is wrong by a much smaller (logarithmic) term. (For the monopartite one, the same guess is correct instead for any $d \geq 1$).

In this section we show nevertheless how to extract a rigorous lower bound from the idea above. We start with the following lemma.

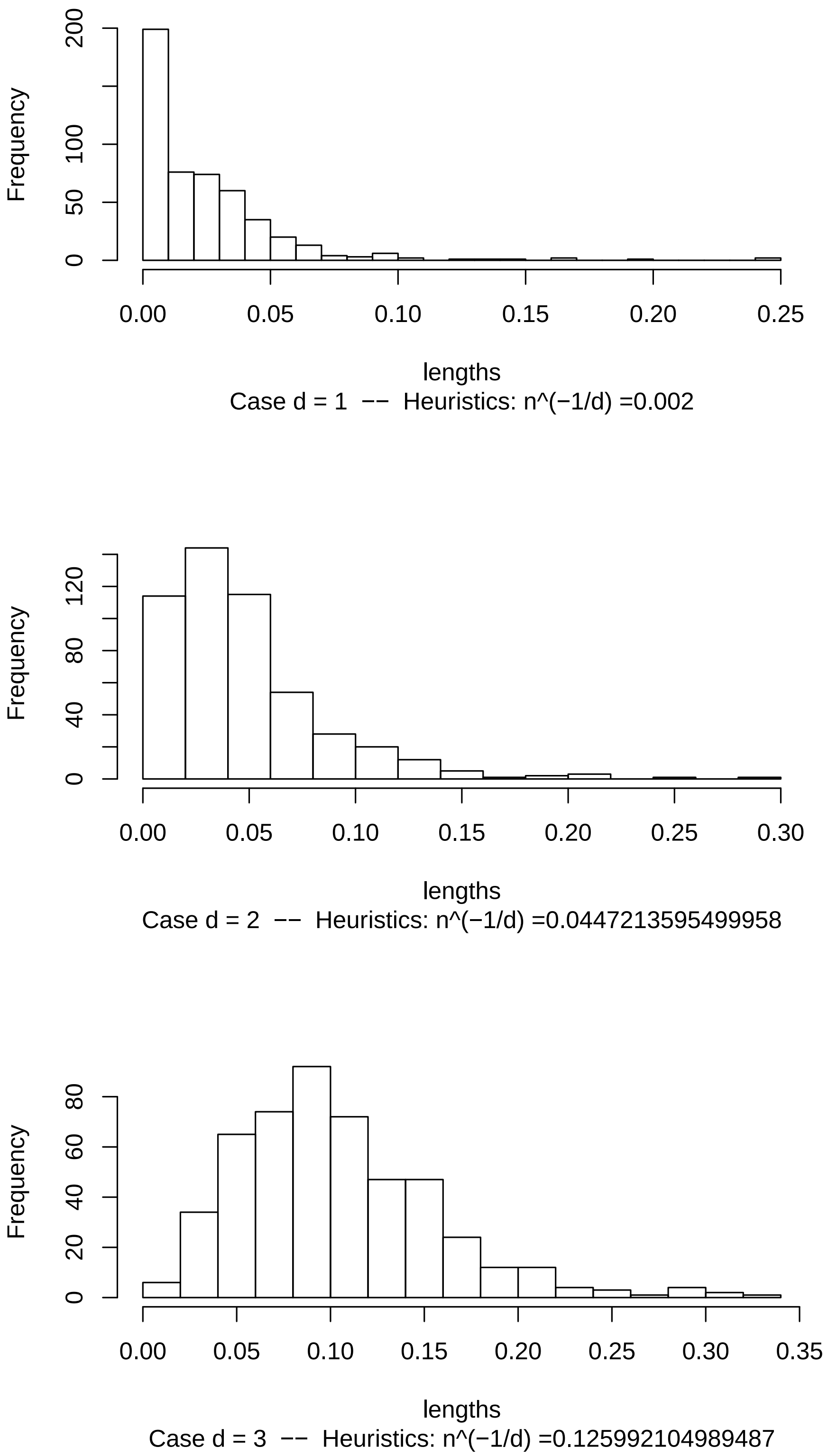
Lemma 2.1. *Let $m \geq 1$ and $(X_i)_{i=1}^m$ i.i.d. on $[0,1]^d$ with uniform distribution. Then, for every $x \in [0,1]^d$,*

$$\mathbb{E}\left[\min_{i=1,\ldots,m} |X_i - x|^p\right] \geq \frac{c(d,p)}{m^{p/d}},$$

where $c(d,p) > 0$ is a positive constant (depending on d and p only).

Proof. As usual with minima of independent random variables (this case $|X_i - x|^p$) it

Figure 3: Histograms of matching lengths with $p = 1$, $n = 500$ and $d \in \{1, 2, 3\}$.



is convenient to consider the *survival* (or reliability) function, i.e., for $t \in [0, \infty)$,

$$
\begin{aligned}
P(\min_{i=1,\ldots,m} |X_i - x|^p > t) &= P(|X_1 - x| > t^{1/p})^m \quad \text{by independence} \\
&= \mathcal{L}^d\left([0,1]^d \setminus \overline{B(x, t^{1/p})}\right)^m \quad \text{since the law is uniform} \\
&\geq (\max\left\{(1 - \omega_d t^{d/p}), 0\right\})^m
\end{aligned}
$$

where we denote $\omega_d = |B(0,1)|$ the measure of the unit ball in $\mathbb{R}^d$. Integrating with respect to t, using the general formula (for non-negative random variables)

$$
\mathbb{E}[Z] = \int_0^\infty P(Z > t) \mathsf{d}t,
$$

gives

$$
\begin{aligned}
\mathbb{E}\left[\min_{i=1,\ldots,m} |X_i - x|^p\right] &= \int_0^\infty P(\min_{i=1,\ldots,m} |X_i - x|^p > t) \mathsf{d}t \\
&\geq \int_0^{\omega_d^{-p/d}} (1 - \omega_d t^{d/p})^m \mathsf{d}t \\
&= \omega_d^{-p/d} \int_0^1 (1 - u^{d/p})^m \mathsf{d}u \\
&\geq \omega_d^{-p/d} \int_0^{m^{-p/d}} (1 - u^{d/p})^m \mathsf{d}u \\
&\geq \omega_d^{-p/d} \left(1 - \frac{1}{m}\right)^m m^{-p/d},
\end{aligned}
$$

hence the thesis with $c(d,p) = \omega_d^{-p/d}/2$ since $(1 - 1/m)^m \geq 1/2$ for every $m \geq 2$. $\square$

Exercise 2.2. Show that the converse inequality holds (possibly with a different $c(d,p)$. For every $m \geq 1$, $x \in [0,1]^d$.

$$
\mathbb{E}\left[\min_{i=1,\ldots,m} |X_i - x|^p\right] \leq \frac{c(d,p)}{m^{p/d}}.
$$

(Hint: following again the same argument, where should be located $x \in [0,1]^d$ so that $\left|[0,1]^d \setminus \overline{B(x, t^{1/p})}\right|$ is maximized?)

As a consequence, we establish the following lower bound.

Proposition 2.3. *There exists $c(d,p) > 0$ such that for every $n \geq 1$*

$$
\mathbb{E}[B_{p,n}] \geq c(d,p) n^{1-p/d}.
$$

Proof. Indeed, given any $\sigma \in \mathcal{S}^n$,

$$
|X_i - Y_{\sigma(i)}|^p \geq \min_{j=1,\ldots,n} |X_i - Y_j|^p,
$$

hence

$$
B_{p,n} \geq \sum_{i=1}^n \min_{j=1,\ldots,n} |X_i - Y_j|^p.
$$

Taking expectation,

$$
\begin{aligned}
\mathbb{E}\left[B_{p,n}\right] &\geq \mathbb{E}\left[\sum_{i=1}^{n} \min_{j=1,\ldots,n} |X_i - Y_j|^p\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\min_{j=1,\ldots,n} |X_i - Y_j|^p\right] \\
&= n\mathbb{E}\left[\min_{j=1,\ldots,n} |X_1 - Y_j|^p\right] \quad \text{the law of } (X_i, Y_1, \ldots, Y_n) \text{ does not depend on } i \\
&= n \int_{[0,1]^d} \mathbb{E}\left[\min_{j=1,\ldots,n} |x - Y_j|^p\right] \mathsf{d}x \quad \text{by independence of } X_1 \text{ and } (Y_j)_j \\
&\geq n \cdot c(d,p) n^{-p/d} \quad \text{by Lemma 2.1.}
\end{aligned}
$$

□

Exercise 2.4. Show that for some $c(d,p) > 0$ one has $\mathbb{E}\left[M_{p,n}\right] \geq c(d,p) n^{1-p/d}$.

3 Concentration Inequalities

In the previous section we provided a lower bound on the expectation of $B_{p,n}$. One may wonder about its variance (or higher moments). In fact, a general phenomenon called *concentration of measure* yields that the random variable $B_{p,n}$ is very close to its average. This is because, in the language of Talagrand [Tal96], the random variable depends in a "regular" way on a large number of independent variables, hence it must be close to a constant (i.e., its expected value). Examples of this phenomenon that fit in this situation and are surely better known are the law of large numbers, where $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}\left[X_i\right]$, or Kolmogorov 0-1 laws, yielding e.g. that $\limsup_{n\to\infty} X_n$ must be a.s. constant, if $(X_n)_n$ are independent real-valued random variables.

There are many (related) results that describe this phenomenon precisely. For our purpose, we use the following inequality. We refer to the already quoted [Tal96, Ste97] but see also [RS13] for another exposition with applications of concentration inequalities in information theory.

Lemma 3.1 (Azuma-Hoeffding-McDiarmid)**.** *Let $(E, \mathcal{E})$ be a measurable space and let $(X_i)_{i=1}^m$ be independent random variables with values in E. Let $f : E^m \to \mathbb{R}$ have bounded differences in the following sense:*

$$
\begin{gathered}
\textit{for every } i \in \{1, \ldots, m\} \textit{ there exists } d_i \geq 0 \textit{ such that} \\
|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_m)| \leq d_i \\
\textit{for every } x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m, x_i' \in E.
\end{gathered}
\tag{3.1}
$$

Then, $Z = f(X_1, \ldots, X_m)$ satisfies

$$
P(|Z - \mathbb{E}\left[Z\right]| > r) \leq 2\exp\left(-\frac{2r^2}{\sum_{i=1}^{m} d_i^2}\right) \quad \textit{for every } r > 0.
$$

In other words, the tail probabilities of Z have a Gaussian-like behaviour, with variance $\sigma^2 = \frac{1}{4}\sum_{i=1}^{m} d_i^2$ (if you have not seen inequalities for Gaussian tail probabilities consider the following exercise).

Exercise 3.2. Let Z be a Gaussian random variable with mean m and variance σ^2. Then

$$
P(|Z - m| > r) \leq \frac{2\sigma}{\sqrt{2\pi} r}\exp\left(-\frac{r^2}{2\sigma^2}\right) \quad \text{for every } r > 0.
$$

(Hint: reduce to a standard case first, $m = 0$, $\sigma^2 = 1$, and then compute by a change of variables

$$\int_r^\infty e^{-x^2/2} \mathsf{d}x = e^{-r^2/2} \int_0^\infty e^{-x^2/2} e^{-xr} \mathsf{d}x \le e^{-r^2/2} \int_0^\infty e^{-xr} \mathsf{d}x,$$

and by symmetry one also deals with the integral between $(-\infty, r)$.)

Remark 3.3. The bounded difference condition is equivalent to a Lipschitz regularity of f with respect to the metric on E given by $\mathsf{d}(x, y) = 1$ if and only if $x \neq y$. Precisely, we have that, for every $i \in \{1, \ldots, m\}$, and $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_m \in E$, the function $x_i \mapsto f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m)$ is Lipschitz with constant bounded by d_i. Other concentration inequalities allow for Lipschitz regularity with respect to different distances (e.g., the Euclidean distance on E if $E \subseteq \mathbb{R}^d$).

As a consequence of the remark above and the fact that pointwise minima of Lipschitz functions (with bounded constants) are still Lipschitz, we have the following result, that can be proved also directly as an exercise.

Exercise 3.4. With the notation of Lemma 3.1, let $(f_u)_{u \in U}$ be a family of non-negative functions on E^m such that the bounded difference condition (3.1) holds with f_u instead of f (and d_i does not depend on $u \in U$). Then, (3.1) holds also for $\inf_{u \in U} f_u$.

As an application, we obtain the following concentration result for $B_{p,n}$.

Proposition 3.5. *For every $r > 0$, one has the inequality*

$$P(|B_{p,n} - \mathbb{E}\left[B_{p,n}\right]| > r) \le 2 \exp\left(-\frac{r^2}{4d^p n}\right),$$

or equivalently (replacing r with $rn^{1-p/d}$),

$$P\left(\left|\frac{B_{p,n}}{n^{1-p/d}} - \frac{\mathbb{E}\left[B_{p,n}\right]}{n^{1-p/d}}\right| > r\right) \le 2 \exp\left(-\frac{r^2 n^{1-2p/d}}{4d^p}\right). \tag{3.2}$$

Proof. Using Exercise 3.4 and the fact that $B_{p,n}$ is defined as a minimum over $\mathcal{S}^n$, it is sufficient to fix $\sigma \in \mathcal{S}^n$ and argue that

$$f_\sigma(x_1, \ldots, x_n, y_1, \ldots, y_n) := \sum_{i=1}^n |x_i - y_{\sigma(i)}|^p$$

satisfies (3.1), and then apply Lemma 3.1 with $E = [0,1]^d$ and $m = 2n$. Indeed, if we choose $i \in \{1, \ldots, 2n\}$ and compute the difference of f_σ when evaluated at two points that may differ only for the component x_i (if $i \le n$) or y_{i-n} (if $i > n$), it is clear that the difference reduces to

$$\left||x_i - y_{\sigma(i)}|^p - |x_i' - y_{\sigma(i)}|^p\right|$$

in the case $i \le n$ and

$$\left||x_{\sigma^{-1}(i-n)} - y_{i-n}|^p - |x_{\sigma^{-1}(i-n)} - y_{i-n}'|^p\right|$$

in the case $i > n$. But in both cases, we trivially estimate each distance between points in $[0,1]^d$ with the diameter of the cube $\sqrt{d}$, so that (3.1) holds with $2d^{p/2}$, and $\sum_{i=1}^{2n} d_i^2 = 8nd^{p/2}$. □

Remark 3.6. In (3.2) the right hand side is infinitesimal (actually summable) as $n \to \infty$ provided that $p < \frac{d}{2}$. An application of Borel-Cantelli lemma gives then the following result:

$$\text{if } p < d/2 \text{ and } \lim_{n\to\infty} \mathbb{E}\left[B_{p,n}\right] n^{1-p/d} = \beta_d \text{ exists (finite), then } \frac{B_{p,n}}{n^{1-p/d}} \to \beta_d \ P\text{-a.s.}$$

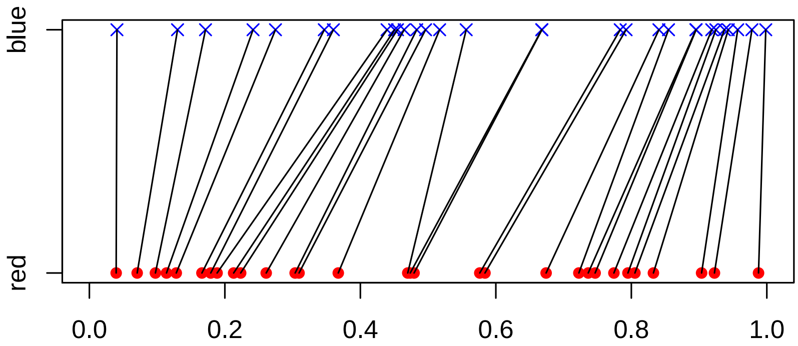


Figure 4: matching of $n = 30$ pairs of i.i.d. uniform points on the unit interval, $p = 1.1$

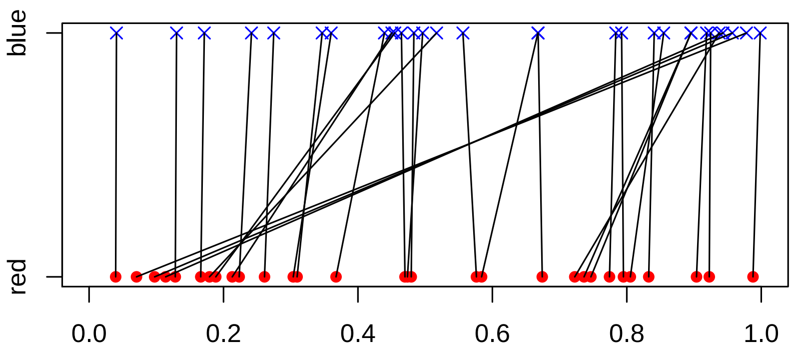


Figure 5: matching of $n = 30$ pairs of i.i.d. uniform points on the unit interval, $p = 0.9$

4 One-dimensional case

When $d = 1$ and $p \geq 1$, the bipartite matching admits a rather simple solution: one has to put in increasing order the red and blue points (separately), and then pair the smallest red with the smallest blue, the second smallest red with the second smallest blue, etc. This can be realized directly from simulations (Figure 4). When $p < 1$, the situation appears to be less “rigid” (Figure 5).

To prove this general fact (which does not depend on the sampling of random points) we introduce the following notation. Given $n \geq 1$ (distinct) points $(z_i)_{i=1}^n$ in $\mathbb{R}$, we denote the k-th smallest value among them by $z_{(k)}$, so that

$$z_{(1)} = \min_i \{z_i\}, \quad z_{(n)} = \max_i \{z_i\},$$

and more generally, for $k \in \{1, \ldots, n-1\}$.

$$z_{(k+1)} = \min \left\{ z_i : z_i > z_{(k)} \right\}.$$

When $z_i = X_i$ or $z_i = Y_i$ are random we write accordingly $X_{(k)}$ and $Y_{(k)}$: these are also known order statistics (of the samples $(X_i)_i$ and $(Y_i)_i$). Notice that we chose for simplicity to define order statistics in the case of distinct points, which happens in the random case P-a.s. (the reader is invited both to give a precise definition in the case of n-points possibly with repetitions, and to argue that P-a.s. all the red and blue points are distinct).

Proposition 4.1 (Matching via order statistics (monotone rearrangement))**.** *Let $n \geq 1$ and let $(x_i)_{i=1}^n$, $(y_i)_{i=1}^n$ be distinct points in $\mathbb{R}$. Then there exists a unique $\sigma^\dagger \in \mathcal{S}^n$ such that*

$$\left\{(x_i, y_{\sigma^\dagger(i)})\right\}_{i=1}^n = \left\{(x_{(i)}, y_{(i)})\right\}_{i=1}^n .$$

Moreover, for every $p \geq 1$,

$$\sum_{i=1}^n |x_i - y_{\sigma^\dagger(i)}|^p = \sum_{i=1}^n |x_{(i)} - y_{(i)}|^p = \min_{\sigma \in \mathcal{S}^n} \sum_{i=1}^n |x_i - y_{\sigma(i)}|^p. \tag{4.1}$$

Proof. The first statement is obvious (all points are assumed to be distinct). For the second part, we argue inductively using the following inequality (which is essentially the case $n = 2$): given $x < x'$, $y < y'$, then

$$|x - y|^p + |x' - y'|^p \leq |x - y'|^p + |x' - y|^p. \tag{4.2}$$

Indeed, assuming that $\bar\sigma$ is a minimizer in (4.1), if $\bar\sigma$ does not match $x_{(1)} := x$ to $y_{(1)} =: y$ but to another point $y' > y_{(1)}$ (and therefore also $y_{(1)}$ is matched to another point $x' > x_{(1)}$) by applying (4.2) with these four points it follows that the matching that pairs $x_{(1)}$ with $y_{(1)}$ and x' with y' and is defined as σ in the other cases has a smaller (or equal) cost. Therefore, we can assume that $\bar\sigma$ matches $x_{(1)}$ with $y_{(1)}$ and removing these points and arguing by induction the thesis would follow.

Of course, it remains to prove (4.2). Without loss of generality, we can assume that $x = 0$ and that $y > 0$, so that it reduces to

$$y^p + |x' - y'|^p \leq (y')^p + |x' - y|^p, \quad \text{for every } x' > 0,\ 0 < y < y',$$

which is equivalent to

$$y^p - |y - x'|^p \leq (y')^p - |y' - x'|^p \quad \text{for every } x' > 0,\ 0 < y < y',$$

i.e., the function $y \mapsto y^p - |y - x'|^p$ is increasing for every $x' > 0$. Differentiating with respect to y gives the expression

$$p(y^{p-1} - \operatorname{sign}(y - x')|y - x'|^{p-1}),$$

which is clearly positive if $y < x'$ while for $y > x'$ is positive since $z \mapsto z^{p-1}$ is increasing (here we use the condition $p \geq 1$). □

Remark 4.2. Notice that if $p > 1$ in (4.2) the inequalities are strict, hence $\sigma^\dagger$ is the unique minimizer in (4.1). For $p = 1$ there may be other minimizers, e.g. if $x_{(1)} \leq x_{(2)} \leq y_{(1)} \leq y_{(2)}$, then pairing $x_{(1)}$ with $y_{(2)}$ and $x_{(2)}$ with $y_{(1)}$ gives the same cost.

Exercise 4.3. Give an explicit example showing that $\sigma^\dagger$ is not necessarily optimal in (4.1) if $0 < p < 1$.

Exercise 4.4. If one associates to $(x_i)_{i=1}^n$ the "empirical" cumulative distribution function $F_x(t) = \frac{1}{n}\sharp\{i : z_i \leq t\}$ and its generalized inverse (*quantile* function), for $\alpha \in (0,1)$,

$$q_x(\alpha) := \inf\{t \in \mathbb{R} : F(t) \geq \alpha\},$$

and similarly for $y = (y_i)_{i=1}^n$, then one has the identity (valid for any $p > 0$)

$$\sum_{i=1}^n |x_{(i)} - y_{(i)}|^p = \int_0^1 |q_x(\alpha) - q_y(\alpha)|^p \mathsf{d}\alpha.$$

Back to the random case, we have now a way to estimate $B_{p,n}$ in terms of the order statistics of $(X_i)_{i=1}^n$ and $(Y_i)_{i=1}^n$. Limiting ourselves to the expectation, we find

$$\mathbb{E}\left[B_{p,n}\right] = \sum_{i=1}^{n} \mathbb{E}\left[|X_{(i)} - Y_{(i)}|^p\right]. \tag{4.3}$$

The variables $X_{(i)}$ and $Y_{(i)}$ are independent, and their law is very well understood, as the following result shows.

Lemma 4.5. *Let $(Z_i)_{i=1}^n$ be i.i.d. random variables uniformly distributed on $[0,1]$. Then for every $t \in [0,1]$, k>1,*

$$P(Z_{(k)} > t > Z_{(k-1)}) = \binom{n}{k-1} t^{k-1}(1-t)^{n-k+1},$$

so that $Z_{(k)}$ has density (on $[0,1]$)

$$f_k(t) = \frac{n!}{(k-1)!(n-k)!} t^{k-1}(1-t)^{n-k}.$$

Proof. The event $Z_{(k)} > t > Z_{(k-1)}$ can be written as the a.s. disjoint union of events where $(k-1)$ random variables are $< t$ and $n-k+1$ are $> t$. Since all these $\binom{n}{k-1}$ events have the same probability, that is

$$P(X_1 < t, X_2 < t, \ldots, X_{k-1} < t, X_k > t, X_{k+1} > t, \ldots X_n > t) = t^{k-1}(1-t)^{n-k+1},$$

the first statement follows. Now we argue by induction: in case $k = 1$, we have

$$P(Z_{(1)} > t) = (1-t)^n,$$

hence $f_1(t) = n(1-t)^{n-1}$. Now assume that we proved the formula for the density of f_{k-1}, and compute

$$\begin{aligned}
-\frac{d}{dt}P(Z_{(k)} > t) &= -\frac{d}{dt}\left[P(Z_{(k)} > t > Z_{(k-1)}) + P(Z_{(k-1)} > t)\right] \\
&= -\binom{n}{k-1}\frac{d}{dt} t^{k-1}(1-t)^{n-k+1} + f_{k-1}(t) \\
&= -\binom{n}{k-1}\left[(k-1)t^{k-2}(1-t)^{n-k+1} - (n-k+1)t^{k-1}(1-t)^{n-k}\right] + \\
&\quad + \frac{n!}{(k-2)!(n-k+1)!} t^{k-2}(1-t)^{n-k+1} \\
&= \frac{n!}{(k-1)!(n-k)!} t^{k-1}(1-t)^{n-k}.
\end{aligned}$$

□

The density f_k belongs to the $\mathrm{Beta}(\alpha, \beta)$ family, which reads

$$\mathrm{Beta}(\alpha, \beta)(t) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1}(1-t)^{\beta-1}$$

for $t \in (0,1)$. In particular f_k is $\mathrm{Beta}(k, n+1-k)$. This allows us to compute all its moments, hence find (in principle) a closed formula for $\mathbb{E}\left[B_{p,n}\right]$, at least when p is even. Let us limit ourselves to the case $p = 2$.

Theorem 4.6. *For $n \geq 1$,*

$$\mathbb{E}\left[B_{2,n}\right] = \frac{1}{3}\frac{n}{n+1}.$$

In particular,

$$\lim_{n\to\infty} \mathbb{E}\left[B_{2,n}\right] = \frac{1}{3}.$$

This shows that the lower bound of Proposition 2.3, that reads in this case

$$\mathbb{E}\left[B_{2,n}\right] \geq c(d,2)n^{-1},$$

misses in fact the correct order, confirming the "educated" guess that the typical matching lengths is of order $n^{-1/2}$. As already anticipated, the situation will be similar also for $d = 2$, but in a subtler way.

Proof. For every $k \in \{1, \dots, n\}$, we have

$$\begin{aligned}
\mathbb{E}\left[|X_{(k)} - Y_{(k)}|^2\right] &= \mathbb{E}\left[X_{(k)}^2\right] - \mathbb{E}\left[X_{(k)}\right]\mathbb{E}\left[Y_{(k)}\right] + \mathbb{E}\left[Y_{(k)}^2\right] \\
&= 2\left(\mathbb{E}\left[X_{(k)}^2\right] - \mathbb{E}\left[X_{(k)}\right]^2\right) \\
&= 2\,\mathrm{Var}\left(X_{(k)}\right).
\end{aligned}$$

Next, we compute

$$\begin{aligned}
\mathbb{E}\left[X_{(k)}\right] &= \int_0^1 t\frac{n!}{(k-1)!(n-k)!}t^{k-1}(1-t)^{n-k}\mathsf{d}t \\
&= \int_0^1 \frac{n!}{(k-1)!(n-k)!}t^{(k+1)-1}(1-t)^{n-k}\mathsf{d}t \\
&= \frac{k}{n+1}\int_0^1 \frac{(n+1)!}{((k+1)-1)!(n-k)!}t^{(k+1)-1}(1-t)^{n-k}\mathsf{d}t \\
&= \frac{k}{n+1},
\end{aligned}$$

where we used the $\mathrm{Beta}(k+1, n+1-k)$ density to argue that the last integral is 1. Similarly,

$$\begin{aligned}
\mathbb{E}\left[X_{(k)}^2\right] &= \int_0^1 t^2\frac{n!}{(k-1)!(n-k)!}t^{k-1}(1-t)^{n-k}\mathsf{d}t \\
&= \int_0^1 \frac{n!}{(k-1)!(n-k)!}t^{(k+2)-1}(1-t)^{n-k}\mathsf{d}t \\
&= \frac{k}{n+1}\int_0^1 \frac{(n+2)!}{((k+2)-1)!(n-k)!}t^{(k+1)-1}(1-t)^{n-k}\mathsf{d}t \\
&= \frac{(k+1)k}{(n+2)(n+1)}.
\end{aligned}$$

Therefore,

$$\mathrm{Var}\left(X_{(k)}\right) = \frac{(k+1)k}{(n+2)(n+1)} - \frac{k^2}{(n+1)^2}.$$

Using (4.3), we obtain

$$\begin{aligned}
\mathbb{E}\left[B_{2,n}\right] &= \sum_{k=1}^{n} 2\left(\frac{(k+1)k}{(n+2)(n+1)} - \frac{k^2}{(n+1)^2}\right) \\
&= 2\left[\frac{n(n+1)(2n+1)}{6(n+2)(n+1)} + \frac{n(n+1)}{2(n+2)(n+1)} - \frac{n(n+1)(2n+1)}{6(n+1)^2}\right] \\
&= \frac{n\left((n+1)(2n+1) + 3(n+1) - (n+2)(2n+1)\right)}{3(n+1)(n+2)} = \frac{1}{3}\frac{n}{n+1}.
\end{aligned}$$

□

Exercise 4.7. Find a closed expression for $\mathbb{E}\left[B_{4,n}\right]$ and show that the following limit holds

$$\lim_{n\to\infty} n\mathbb{E}\left[B_{4,n}\right] = \frac{2}{5}.$$

Exercise 4.8. Consider the random Euclidean bipartite matching when $(X_i)_{i=1}^n$, $(Y_i)_{i=1}^n$ are i.i.d. with exponential law of parameter 1. Is it possible to find closed expressions for $\mathbb{E}[B_{2,n}]$, or at least provide precise asymptotics as $n \to \infty$? It may be useful to use the fact that $X_{(k)}$ has the same law as $\sum_{i=1}^k \frac{Z_i}{n-i+1}$ where Z_i are i.i.d. with exponential law of parameter 1.

5 Self-averaging argument

In this section we describe an approach [BdMM02, BB13] that ultimately yields existence of

$$\lim_{n\to\infty} \mathbb{E}[B_{p,n}]\, n^{-1+p/d} \in (0,\infty),$$

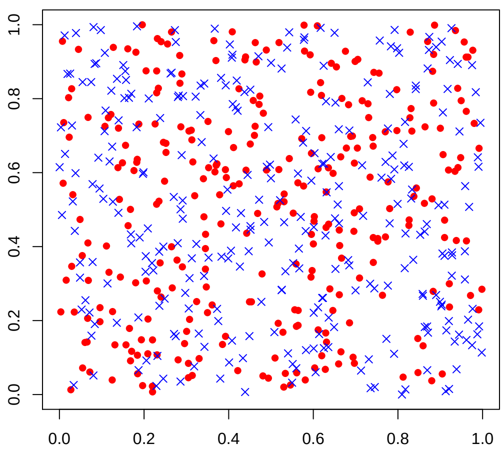
provided that $0 < p < d/2$. The main idea is that, by decomposing the original cube into smaller cubes, one obtains a similar picture (only with less points). Hence, one could construct a matching by gluing together "local" matchings on each small cube. Since the total cost of the optimal matching must be smaller, this would lead to a monotonicity property of a suitably rescaled $B_{p,n}$ (it should be decreasing) and therefore existence of a limit (simply because any decreasing sequence converges to its infimum).

There are at least two issues to be solved in order to put this idea into a rigorous formulation.

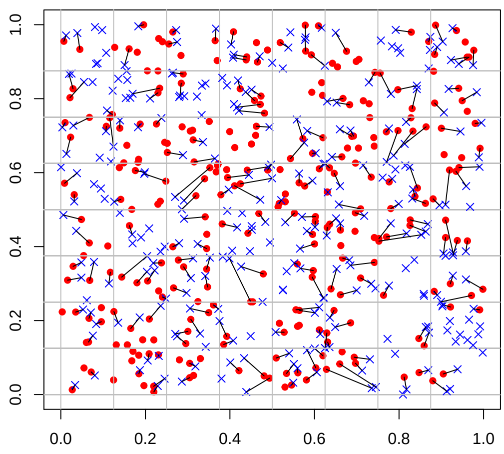
The first one concerns the fact that, if we choose any (good) region $U \subseteq [0,1]^d$, we may observe a different number of red and blue points. How to define a "local" matching in this case? We choose to extend the matching problem in the following way: if we observe that there are $n(X;U)$ red points and $n(Y;U)$ blue points, we try to match (in an optimal way), only $\min\{n(X;U), n(Y;U)\}$ points and leave the remaining $|n(X;U) - n(Y;U)|$ points "unmatched". This will add an extra error term in the monotonicity argument above, since we will need to pair the leftover points outside the small cubes.

The main idea will be then to iterate this construction going from smaller to larger scales, e.g. in a dyadic way (Figure 6).

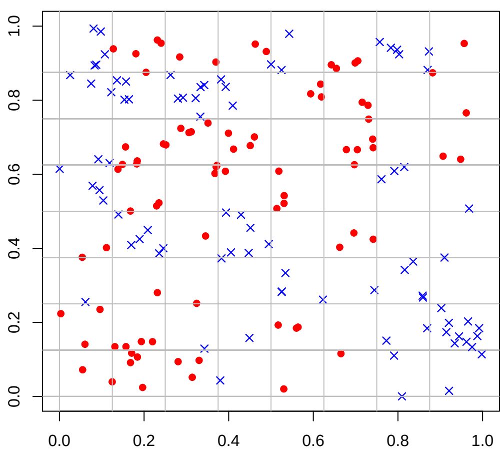
Figure 6: Defining a matching between $n = 300$ pair of points by iteratively "gluing local matchings at scales 2^{-k} with $k = 3, 2, 1, 0$ ($p = 1$).



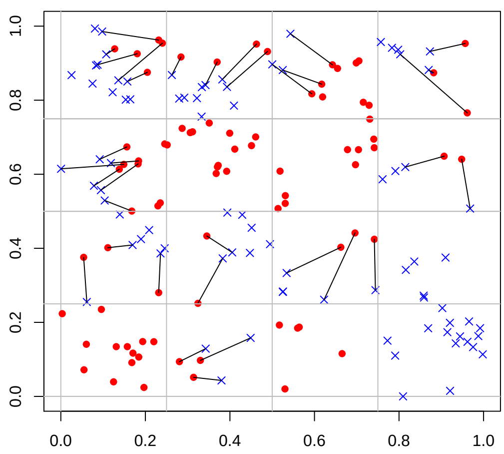
(a) $n = 300$ pairs of i.i.d. uniform points.



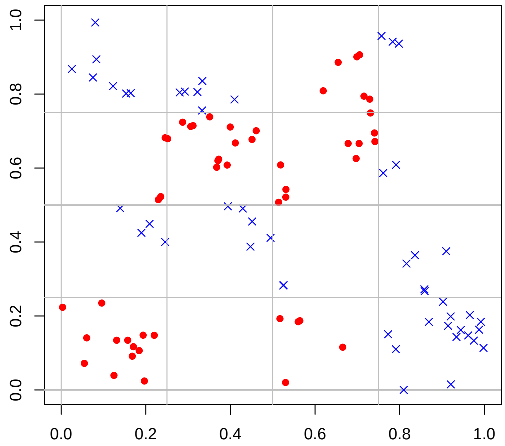
(b) Matchings on squares of side lengths 2^{-3}.



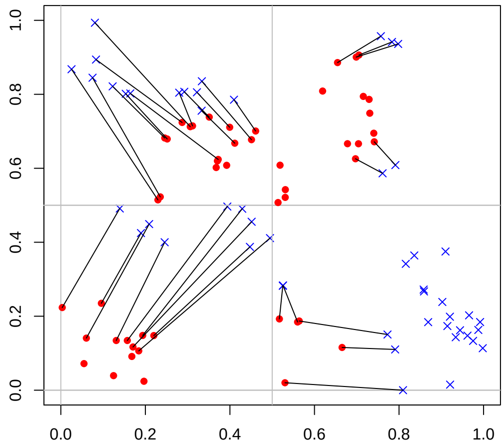
(c) Unmatched points in the previous step.



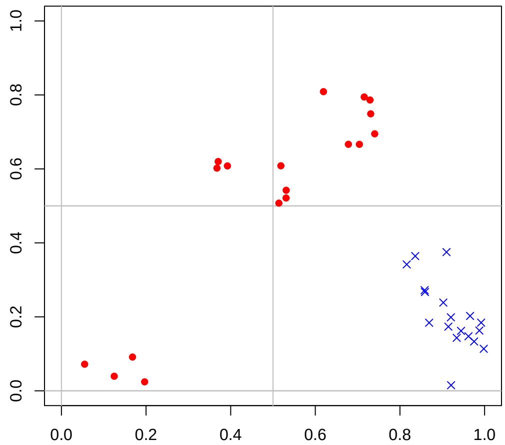
(d) Matchings on squares of side lengths 2^{-2}.



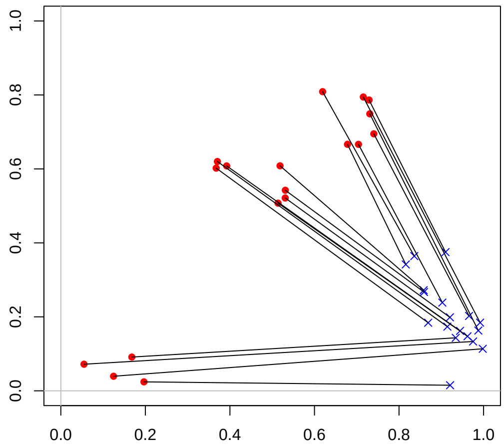
(e) Unmatched points in the previous step.



(f) matchings on squares of side lengths 2^{-1}



(g) Unmatched points in the previous step.



(h) Matching on unit square.

Let us introduce some notation in order do properly define this extension. Given $n \in \mathbb{N}$, $x = (x_i)_{i=1}^n \subseteq \mathbb{R}^d$, write

$$\mathcal{I}(x; U) := \{ i \in \{1, \ldots, n\} \, : \, x_i \in U \} \, ,$$

for the subset (of indices) of points in U, so that $n(x; U) := \sharp \mathcal{I}(x; U)$. If we are given also $m \in \mathbb{N}$ points $y = (y_i)_{i=1}^m \subseteq \mathbb{R}^d$, the "local" matchings in U will be parametrized by relations $\sigma \subseteq \mathcal{I}(x; U) \times \mathcal{I}(y; U)$ that are injective functions from $\mathcal{I}(x; U)$ to $\mathcal{I}(y; U)$ if $n(x; U) \leq n(y; U)$ or viceversa if $n(x; U) \geq n(y; U)$. Let us write $\mathcal{S}(x, y; U)$ for such a set and finally let

$$b_p(x, y; U) := \min_{\sigma \in \mathcal{S}(x,y;U)} \sum_{(i,j) \in \sigma} |x_i - y_j|^p.$$

for the "local" bipartite matching cost in U. Notice that there may be no (red or blue) points in U, and in that case the local cost is zero.

We will need only basic properties of b_u, summarized in the following lemma.

Lemma 5.1. *Let n, $m \in \mathbb{N}$ $x = (x_i)_{i=1}^n$, $y = (y_i)_{i=1}^m \subseteq \mathbb{R}^d$, $U \subseteq \mathbb{R}^d$ Borel. Then,*

1. *b_p is local, i.e., $b_p(x, y; U) = b_p((x_i)_{i \in \mathcal{I}(x;U)}, (y_i)_{i \in \mathcal{I}(y;U)}, U)$.*
2. *b_p is translation invariant: if $v \in \mathbb{R}^d$, then*
$$b_p(x + v, y + v; U + v) = b_p(x, y; U),$$
where $x + v = (x_i + v)_{i=1}^n$, $y + v = (y_i + v)_{i=1}^m$, $U + v = \{u + v : u \in U\}$.
3. *b_p is p-homogeneous with respect to dilations: if $\lambda \in (0, \infty)$, then*
$$b_p(\lambda x, \lambda y, \lambda U) = \lambda^p b_p(x, y; U),$$
where $\lambda x = (\lambda x_i)_{i=1}^n$, $\lambda y = (\lambda y_i)_{i=1}^m$, $\lambda U = \{\lambda u : u \in U\}$.
4. *b_p is p-subadditive: for every (Borel) partition $U = \bigcup_{k=1}^K U_k$, one has*
$$b_p(x, y; U) \leq \sum_{k=1}^K b_p(x, y; U_k) + |n(x; U_k) - n(y; U_k)| \operatorname{diam}(U)^p,$$
where $\operatorname{diam}(U) = \sup\{|x - y| \,:\, x, y \in U)\}$.

Proof. We leave properties *1-3* to the reader. To show *4*, notice that any collection of matchings $(\sigma_k)_{k=1}^K$ with $\sigma_k \in \mathcal{S}(x, y; U_k)$ gives a relation $\sigma := \bigcup_{k=1}^K \sigma_k$ that is left and right injective (i.e., if $(i, j) \in \sigma$, there are no other pairs which share at least one coordinate with (i, j)). To obtain an injective function, it is sufficient to arbitrarily match points from the remaining ones, up to $\min\{n(x; U), n(y; U)\}$. Each will give a contribution smaller than $\operatorname{diam}(U)^p$, and the overall number is bounded from above by $\sum_{k=1}^K |n(x; U_k) - n(y; U_k)|$. □

Next, we address the second issue: when $X = (X_i)_{i=1}^n$, $Y = (Y_i)_{i=1}^n$ are random i.i.d. uniform on the cube $[0, 1]^d$, the probabilistic properties of the points inside U are different than the original ones (even if U is a sub-cube). For example (exercise)

$$n(X; U)$$

has a Binomial law with parameters $(n, |U|)$. In order to gain self-similarity (at least in law), the idea here is to pick n (the initial number of points) to be also *random*, with a Poisson law, with intensity λ, i.e., we impose that

$$P(n(X; [0, 1]^d) = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{for } k \in \mathbb{N}.$$

Exercise 5.2. Show that a Poisson random variable Z with parameter λ has mean λ and variance λ and that, for $\alpha \in (0, 1]$, $\lambda \geq 1$,

$$\mathbb{E}\left[Z^\alpha\right] \geq c(\alpha)\lambda^\alpha,$$

where $c(\alpha) > 0$ depends on α only. *(Hint: use the inequality $Z^\alpha \geq \lambda^\alpha - |Z - \lambda|^\alpha$)*

This *Poissonization* procedure can be rigorously constructed as follows. Let $\lambda > 0$, let N_X and N_Y be independent Poisson random variables with intensity λ. Let also $(X_i)_{i=1}^\infty$, $(Y_i)_{i=1}^\infty$ be i.i.d. uniform random variables in the cube, also independent of N_X and N_Y. We consider then the two independent *Poisson point processes of intensity* λ on $[0, 1]^d$ defined as

$$X := (X_i)_{i=1}^{N_X}, \quad Y := (Y_i)_{i=1}^{N_Y},$$

i.e., on the event $\{N_X = k\}$ one has $X = (X_i)_{i=1}^k$ (and similarly for Y).

We collect some useful properties of this construction in the following lemma.

Lemma 5.3. *Let $\lambda > 0$ and let $X = (X_i)_{i=1}^{N_X}$ be a Poisson point process of intensity λ on $[0,1]^d$. For every $U \subseteq [0,1]^d$ Borel with $|U| > 0$, the random variable*

$$n(X;U) = \sharp\mathcal{I}(X;U)$$

has Poisson law of parameter $\lambda|U|$ and, conditionally to the event $\{n(X;U) = k\}$, the k random variables $(X_i)_{i\in\mathcal{I}(X;U)}$ are independent with uniform law on U (i.e., with density $\frac{1}{|U|}$ on U).

Proof. Given $k \geq 1$ and Borel subsets $(A_i)_{i=1}^k$ of $[0,1]^d$, we write

$$I = \{\sharp\mathcal{I}(X;U) = k\},$$

$$A = \left\{\sharp\mathcal{I}(X;U) = k, \quad (X_i)_{i\in\mathcal{I}(X;U)} \in \prod_{i=1}^k A_i\right\},$$

where the notation $(X_i)_{i\in\mathcal{I}(X;U)} \in \prod_{i=1}^k A_i$ means that the k points $(X_i)_{i\in\mathcal{I}(X;U)}$ are imposed to belong respectively to the k sets $(A_1, A_2, \ldots, A_k)$ keeping the natural order in their parametrization (i.e., the point X_i with smallest index $i \in \mathcal{I}(X;U)$ has to belong to A_1, and so on). The thesis can be stated as

$$P(I) = \frac{(\lambda|U|)^k}{k!} e^{-\lambda|U|}, \tag{5.1}$$

$$P(A|I) = \prod_{i=1}^k \frac{|A_i|}{|U|}. \tag{5.2}$$

We compute separately $P(I)$ and $P(A \cap I) = P(A)$. In both cases, we decompose over the alternatives $\{N_X = n\}$ (clearly only with $n \geq k$), so that (with $B = I$ or $B = A \cap I$),

$$P(B) = \sum_{n=k}^{+\infty} P(B|N_X = n)P(N_X = n).$$

By assumption,

$$P(N_X = n) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

To compute $P(I|N_X = n)$, it is sufficient to notice that when $X = (X_i)_{i=1}^n$ are i.i.d. uniform points then

$$n(X;U) = \sum_{i=1}^n I_U(X_i)$$

is the sum of n independent Bernoulli random variables with parameter $|U|$, hence it has Binomial law with parameters $(n, \mathcal{L}^d(U))$, so that

$$P(I|N_X = n) = \binom{n}{k} |U|^k (1 - |U|)^{n-k}.$$

Therefore,

$$\begin{aligned} P(I) &= \sum_{n=k}^{+\infty} \binom{n}{k} |U|^k (1-|U|)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \frac{(\lambda|U|)^k}{k!} e^{-\lambda} \sum_{n=k}^{\infty} \frac{\lambda^{n-k}(1-|U|)^{n-k}}{(n-k)!} \\ &= \frac{(\lambda|U|)^k}{k!} e^{-\lambda|U|}. \end{aligned}$$

This proves (5.1). To show (5.2), we notice that the event $A \cap \{N_X = n\}$ can be decomposed into the disjoint union of events of the type

$$\bigcap_{j=1}^{n} \{X_j \in B_j\},$$

where each B_j is $[0,1]^d \setminus U$ or one of the sets $(A_i)_{i=1}^k$. Moreover, each set A_i must appear only once and in the given order. Clearly, each event of the type above has probability

$$P(\bigcap_{j=1}^{n} \{X_j \in B_j\} | N_X = n) = \prod_{j=1}^{n} |B_j| = (1-|U|)^{n-k} \prod_{i=1}^{k} |A_i|,$$

Their number is $\binom{n}{k}$, for once we specify the subset of k indices that are not $[0,1]^d \setminus U$ the sequence is uniquely determined because the sets $(A_i)_{i=1}^n$ appear in a specified order. Hence,

$$P(A|N_X = n) = \binom{n}{k}(1-|U|)^{n-k} \prod_{i=1}^{k} |A_i|.$$

The computation now is identical to the previous case, with $|U|^k$ replaced by $\prod_{i=1}^k |A_i|$. We conclude that

$$P(A) = \frac{\lambda^k \prod_{i=1}^k |A_i|}{k!} e^{-\lambda|U|}.$$

Taking the quotient, (5.2) follows. □

As a consequence, we obtain the following *self-similarity* property of X.

Corollary 5.4. *With the notation above, let also $Q = [a,b]^d \subseteq [0,1]^d$.*[1] *Then the process obtained by "restriction" of X to Q and transformation to $[0,1]^d$ via an affine map, i.e.,*

$$((X_i - a)/(b-a))_{i \in \mathcal{I}(X;Q)},$$

is a Poisson point process on $[0,1]^d$ of intensity $\lambda(b-a)^d$.

Proof. Indeed, it is sufficient to let $U = Q$ in the result above and use the fact that i.i.d. random variables with uniform law on Q are transformed to i.i.d. random variables with uniform law on $[0,1]^d$ via the affine map. □

As we mentioned at the beginning of this section, we will not establish an exact monotonicity when decomposing into smaller cubes. To take care of this fact and still show existence of a limit, we need the following analytical result.

Lemma 5.5. *Let $\alpha > 0$, $c \geq 0$, $f : [1,\infty) \to [0,\infty)$ be continuous and such that, for every $m \geq 1$ integer, $\lambda \geq 1$ real, one has*

$$f(m\lambda) \leq f(\lambda) + c\lambda^{-\alpha}. \tag{5.3}$$

Then $\lim_{\lambda\to\infty} f(\lambda)$ exists.

Proof. We crucially use the following fact: for any open interval $(a,b) \subseteq \mathbb{R}$, the union $\bigcup_{m=1}^{\infty}(ma, mb)$ contains a half-line $(A, +\infty)$ for some $A > 0$. Indeed, one has $(ma, mb) \cap ((m+1)a, (m+1)b) \neq \emptyset$ if $mb > (m+1)a$, which holds for every $m > a/(b-a)$.

[1] In fact, I should always consider here and in what follows more general cubes of the type $Q = \{v + x : x \in [0,(b-a)]^p\}$, where $v \in \mathbb{R}^d$, $a, b \in \mathbb{R}$, $a < b$ and $Q \subseteq [0,1]^d$, but I will not change it for simplicity of notation in the exposition.

Since both f and $c\lambda^{-\alpha}$ are continuous on $[1, 2]$, they are bounded and, by (5.3), for every $m \geq 1$, $\lambda \in [1, 2]$,

$$f(m\lambda) \leq \sup_{s \in [1,2]} \left(f(s) + cs^{-\alpha} \right) < \infty.$$

since $\bigcup_{m=1}^{\infty} [m, 2m] = [1, \infty)$, it follows that f is uniformly bounded on $[1, \infty)$. To show that the limit exists, it is sufficient to argue that

$$\limsup_{\lambda \to \infty} f(\lambda) \leq \liminf_{\lambda \to \infty} f(\lambda).$$

Let $\varepsilon > 0$ and $x_0 > 0$ be such that $cx^{-\alpha} < \varepsilon$ for every $x > x_0$. Let also $\lambda_0 \in (x_0, \infty)$ be such that

$$f(\lambda_0) < \liminf_{\lambda \to \infty} f(\lambda) + \varepsilon.$$

By continuity of f, there exists (a, b) such that the same inequality holds on it (and, without loss of generality, $a > x_0$). For every $m \geq 1$, $x \in I$ it follows that

$$\begin{aligned} f(mx) \leq f(x) + cx^{-\alpha} &< \liminf_{\lambda \to \infty} f(\lambda) + \varepsilon + cx^{-\alpha} \\ &\leq \liminf_{\lambda \to \infty} f(\lambda) + 2\varepsilon. \end{aligned}$$

By the fact established at the beginning of the proof shows that, it follows that

$$f(x) \leq \liminf_{\lambda \to \infty} f(\lambda) + 2\varepsilon$$

for every $x \in (A, +\infty)$, for some $A > 0$. It follows that

$$\limsup_{\lambda \to \infty} f(\lambda) \leq \liminf_{\lambda \to \infty} f(\lambda) + 2\varepsilon,$$

hence the thesis letting $\varepsilon \to 0$. $\square$

Theorem 5.6. *Let $p \in (0, d/2)$ and let X_λ, Y_λ be independent Poisson point processes on $[0, 1]^d$ with intensity λ. Then*

$$\lim_{\lambda \to \infty} \frac{\mathbb{E}\left[b_p(X_\lambda, Y_\lambda, [0, 1]^d) \right]}{\lambda^{1-p/d}} \in (0, \infty)$$

exists.

Proof. Fix λ and write $f(\lambda) = \mathbb{E}\left[b_p(X, Y; [0, 1]^d) \right] \lambda^{p-d}$, for X, Y independent Poisson point processes on $[0, 1]^d$ with intensity λ^d (notice the slight change of parametrization). The thesis becomes then that the limit $\lim_{\lambda \to \infty} f(\lambda)$ exists finite and strictly positive. Let us notice that, by conditioning upon the event $\{N_X = n, N_Y = m\}$, we can write

$$f(\lambda) = \lambda^{p-d} \sum_{n,m=1}^{\infty} \mathbb{E}\left[b_p((X_i)_{i=1}^n, (Y_j)_{j=1}^m; [0, 1]^d) \right] \frac{\lambda^{nd}}{n!} e^{-\lambda^d} \frac{\lambda^{md}}{m!} e^{-\lambda^d}. \tag{5.4}$$

where we used also that, if one among n, m is null, then the matching cost is by definition zero. We notice that $\lambda \mapsto f(\lambda)$ is a continuous function, e.g. by Lebesgue dominated convergence (use the trivial inequality $\mathbb{E}\left[b_{p,n}((X_i)_{i=1}^n, (Y_j)_{j=1}^m \right] \leq d^{p/2}(n+m)$). Hence, we only need to show (5.3) so that Lemma 5.5 applies. Before we do so, we also argue that $f(\lambda) > c(d, p) > 0$ for some constant independent of λ.[2] This amounts to go back to Proposition 2.3 and extend the result to the Poisson case. Indeed, if both n, m are

[2]There are surely shorter ways to prove this, but all the proofs I found were less elementary. If you find a shorter and more elementary I would like to know.

natural and positive, we suggest as an instructive exercise to check that Proposition 2.3 gives the inequality

$$\mathbb{E}\left[b_p((X_i)_{i=1}^n, (Y_j)_{j=1}^m; [0,1]^d)\right] \geq c(d,p)\frac{\min\{n,m\}}{\max\{n,m\}^{p/d}}.$$

Therefore, we can write

$$f(\lambda) \geq \lambda^{p-d} \sum_{n,m=1}^{\infty} c(d,p) n^{1-p/d} \frac{\lambda^{nd}}{n!} e^{-\lambda^d} \geq c(d,p)\lambda^{p-d}\mathbb{E}\left[\frac{\min\{N_X, N_Y\}}{\max\{N_X, N_Y\}^{p/d}}\right].$$

To obtain a lower bound, we try to argue as in Exercise 5.2 with $\alpha \in (0,1]$, in place of Z Poisson, the variables $Z_1 = \min\{N_X, N_Y\}$ and $Z_2 = \max\{N_X, N_Y\}$. Indeed we have

$$\begin{aligned}
&\max\left\{\mathbb{E}\left[\left|Z_1 - \lambda^d\right|\right], \mathbb{E}\left[\left|Z_2 - \lambda^d\right|\right]\right\} \leq \\
&\mathbb{E}\left[\left|\min(N_X, N_Y) - \lambda^d\right|\right] + \mathbb{E}\left[\left|\max(N_X, N_Y) - \lambda^d\right|\right] \\
&= \mathbb{E}\left[\left|\min(N_X, N_Y) - \lambda^d\right| + \left|\max(N_X, N_Y) - \lambda^d\right|\right] \\
&= \mathbb{E}\left[\left|N_X - \lambda^d\right| + \left|N_Y - \lambda^d\right|\right] \leq 2\lambda^{d/2},
\end{aligned}$$

If $\alpha \leq 1$, the triangle inequality holds, hence $\lambda^{\alpha d} \leq |Z|^\alpha + |Z - \lambda^d|^\alpha$ and taking expectation we have

$$\begin{aligned}
\mathbb{E}\left[Z_1^\alpha\right] &\geq \lambda^{\alpha d} - \mathbb{E}\left[|Z_1 - \lambda^d|^\alpha\right] \\
&\geq \lambda^{\alpha d} - 2^\alpha \lambda^{\alpha d/2} \geq c(d,\alpha)\lambda^{\alpha d} \text{ for } \lambda \text{ large enough,}
\end{aligned}$$

having used Hölder inequality with exponent $1/\alpha \geq 1$ to estimate from above

$$\mathbb{E}\left[|Z_1 - \lambda^d|^\alpha\right] \leq \mathbb{E}\left[|Z_1 - \lambda^d|\right]^\alpha \leq 2^\alpha \lambda^{d\alpha/2}.$$

For Z_2, we need to argue in the opposite direction, so that

$$\mathbb{E}\left[Z_2\right] \leq \lambda^d + \mathbb{E}\left[|Z_2 - \lambda^d|\right] \leq \lambda^d + 2\lambda^{d/2} \leq c'(1,d)\lambda^d$$

for λ large enough. To conclude, we use the fact that, letting $\gamma = p/d \in (0,1/2)$ and $\beta = 1 + \gamma > 1$, the function $h(x,y) = \frac{x^\beta}{y^\gamma}$ is convex on $(0,+\infty) \times (0,+\infty)$. Indeed, its Hessian matrix is given by the symmetric matrix

$$\begin{pmatrix} \frac{\partial^2}{\partial x^2} h & \frac{\partial^2}{\partial x \partial y} h \\ \frac{\partial^2}{\partial x \partial y} h & \frac{\partial^2}{\partial y^2} h \end{pmatrix} = \begin{pmatrix} \beta(\beta-1)x^{\beta-2}y^{-\gamma} & -\beta\gamma x^{\beta-1}y^{-\gamma-1} \\ -\beta\gamma x^{\beta-1}y^{-\gamma-1} & \gamma(\gamma+1)x^\beta y^{-\gamma-2} \end{pmatrix}$$

that has positive trace and determinant equal to

$$\beta\gamma\left[(\beta-1)(\gamma+1) - \beta\gamma\right] x^{2\beta-2}y^{-2\gamma-2} = 0,$$

Therefore, by Jensen inequality

$$\begin{aligned}
\mathbb{E}\left[\frac{Z_1}{Z_2^{p/d}}\right] &= \mathbb{E}\left[h\left(Z_1^{1/\beta}, Z_2\right)\right] \geq h\left(\mathbb{E}\left[Z_1^{1/\beta}\right], \mathbb{E}\left[Z_2\right]\right) \\
&= \frac{\mathbb{E}\left[Z_1^{1/\beta}\right]^\beta}{\mathbb{E}\left[Z_2\right]^{p/d}} \\
&\geq \frac{c(\beta,d)^\beta \lambda^d}{c'(1,d)^{p/d}\lambda^p} \\
&\geq c''(p,d)\lambda^{d-p},
\end{aligned}$$

that gives $f(\lambda) > c''(p,d) > 0$ for λ large enough.

Let us now focus on the monotonicity property (5.3). If $Q = [a,b]^d$, then

$$\begin{aligned}
\mathbb{E}\left[b_p(X,Y;Q)\right] &= \mathbb{E}\left[b_p((X_i)_{i\in\mathcal{I}(X;Q)},(Y_i)_{i\in\mathcal{I}(Y;Q)};Q)\right] \quad \text{by Lemma 5.1.}\mathit{1}\\
&= (b-a)^p\mathbb{E}\left[b_p\left(((X_i-a)/(b-a))_{i\in\mathcal{I}(X;Q)},((Y_i-a)/(b-a))_{i\in\mathcal{I}(Y;Q)};[0,1]^d\right)\right]\\
&\qquad \text{by Lemma 5.1.}\mathit{2}\text{–}\mathit{3}\\
&= (b-a)^p\mathbb{E}\left[b_p(\tilde{X},\tilde{Y};[0,1]^d)\right]\\
&\qquad \text{by Corollary 5.4, where } \tilde{X} \text{ and } \tilde{Y} \text{ have now intensity } (\lambda(b-a))^d.
\end{aligned}$$

Multiplying both sides by λ^{p-d}, it follows that

$$\lambda^{p-d}\mathbb{E}\left[b_p(X,Y;Q)\right] = (b-a)^d f(\lambda(b-a)). \tag{5.5}$$

Let now $m \geq 1$ be an integer. Decomposing $[0,1]^d$ into m^d disjoint subcubes $(Q_i)_{i=1}^{m^d}$ of side length $1/m$, it follows from Lemma 5.1.*4* and the identity just obtained that

$$\begin{aligned}
f(\lambda) &\leq \lambda^{p-d}\sum_{i=1}^{m^d}\mathbb{E}\left[b_p(X,Y;Q_i)\right] + \mathbb{E}\left[|n(X;Q_i)-n(Y;Q_i)|\right]\\
&= \sum_{i=1}^{m^d} m^{-d} f(\lambda/m) + \lambda^{p-d}\mathbb{E}\left[|n(X;Q_i)-n(Y;Q_i)|\right]\\
&= f(\lambda/m) + \lambda^{p-d}\sum_{k=1}^{m^d}\mathbb{E}\left[|n(X;Q_i)-n(Y;Q_i)|\right].
\end{aligned}$$

To estimate the last sum, we use the first part of Lemma 5.3, so that the random variables $n(X;Q_k)$ and $n(Y;Q_k)$ are independent Poisson with parameter $\lambda|Q_k| = \lambda^d m^{-d}$, hence

$$\begin{aligned}
\mathbb{E}\left[|n(X;Q_i)-n(Y;Q_i)|\right] &\leq \mathbb{E}\left[|n(X;Q_i)-n(Y;Q_i)|^2\right]^{1/2}\\
&= \left(2\,\mathrm{Var}\left(n(X;Q_i)\right)\right)^{1/2}\\
&= \sqrt{2}\lambda^{d/2}m^{-d/2}.
\end{aligned}$$

Summing all the m^d contributions, we conclude that

$$f(\lambda) \leq f(\lambda/m) + \sqrt{2}\lambda^{p-d/2}m^{d/2}$$

or equivalently, replacing λ with λ/m,

$$f(m\lambda) \leq f(\lambda) + \sqrt{2}m^p\lambda^{p-d/2}. \tag{5.6}$$

This is almost (5.3) with $\alpha = d/2 - p > 0$, but the constant c depends on m and diverges for $m \to \infty$. To solve this issue, we need to argue using "multiple scales" instead of decomposing directly $[0,1]^d$ into m^d sub-cubes. Notice indeed, that iterating k times (5.6) with $m = 2$ gives

$$\begin{aligned}
f(2^k\lambda) &\leq f(2^{k-1}\lambda) + \sqrt{2}2^p\lambda^{-\alpha}2^{-(k-1)\alpha}\\
&\leq f(2^{k-2}\lambda) + \sqrt{2}2^p\lambda^{-\alpha}(2^{-(k-1)\alpha} + 2^{-(k-2)\alpha})\\
&\leq \ldots \leq f(\lambda) + \sqrt{2}2^p\lambda^{-\alpha}\sum_{i=0}^{k-1}2^{-\alpha i}\\
&\leq f(\lambda) + c\lambda^{-\alpha},
\end{aligned} \tag{5.7}$$

with $c = \sqrt{2}2^p\sum_{i=0}^{+\infty}2^{-\alpha i}$.

To argue similarly, for a general $m > 1$, we need however to repeat all the derivation of (5.6) in a more careful way. Choose $k \geq 1$ so that $2^{k-1} < m \leq 2^k$ and consider the cube $\tilde{Q} = [0, 2^k/m]^d \supseteq [0,1]^d$. Notice that

$$\lambda^{p-d}\mathbb{E}\left[b_p(X,Y;\tilde{Q})\right] = \lambda^{p-d}\mathbb{E}\left[b_p(X,Y;Q)\right],$$

since the point processes X and Y take values in Q. For every $i \in \{1, \ldots, k\}$, we decompose $\tilde{Q}$ into 2^{id} sub-cubes $(Q_{i,j})_{j=1}^{2^{id}}$ of side length $2^{k-i}/m$. When $i = k$, this decomposition induces one of $[0,1]^d$ into m^d sub-cubes. However, not all cubes $Q_{i,j}$ will be entirely contained in $[0,1]^d$, hence the self-similarity identity (5.5) fails for them. When $i = k$ we can however argue that if a sub-cube $\tilde{Q}_i$ is disjoint from $[0,1]^d$, then $b_p(X,Y;\tilde{Q}_i) = 0$. On the other side, each term of the type $\mathbb{E}\left[|n(X;Q_{i,j}) - n(Y;Q_{i,j})|\right]$ can be always bounded from above by $\sqrt{2}\left(\lambda 2^{(k-i)}/m\right)^{d/2}$, since $n(X;Q_{i,j})$ and $n(Y;Q_{i,j})$ are independent Poisson random variables with a (possibly smaller) parameter $\lambda^d|Q_{i,j} \cap [0,1]^d|$. Finally, to mimic (5.7), we use the fact that the decomposition can be thought as an iterative procedure, hence we can apply Lemma 5.1.*4* to each $Q_{i,j}$, that has diameter $\sqrt{d}2^{k-i}/m$, and summing upon j (with i fixed) obtaining

$$\begin{aligned}\sum_{j=1}^{2^{id}} \mathbb{E}\left[b_p(X,Y;Q_{i,j})\right] \leq &\sum_{j=1}^{2^{(i+1)d}} \mathbb{E}\left[b_p(X,Y;Q_{i+1,j})\right] \\ &+ \left(\sqrt{d}2^{k-i}/m\right)^p \mathbb{E}\left[|n(X;Q_{i,j}) - n(Y;Q_{i,j})|\right].\end{aligned}$$

Let us focus on the "error terms"

$$\begin{aligned}\sum_{j=1}^{2^{id}} \sqrt{d}\left(2^{k-i}/m\right)^p \mathbb{E}\left[|n(X;Q_{i,j}) - n(Y;Q_{i,j})|\right] &\leq 2^{id}\left(\sqrt{d}2^{k-i}/m\right)^p \sqrt{2}\left(\lambda 2^{(k-i)}/m\right)^{d/2} \\ &= \sqrt{2d^p}\lambda^{d/2}2^{i(d/2-p)},\end{aligned}$$

having used that $m/2^k \leq 1$.

Multiplying by λ^{p-d} and iterating this procedure from $i = 0$ to $i = k-1$ leads to

$$f(\lambda) \leq \left(\sum_{j=1}^{2^{kd}} \lambda^{p-d}\mathbb{E}\left[b_p(X,Y;Q_{k,j}\right]\right) + \sqrt{2d^p}\lambda^{p-d/2}\sum_{i=0}^{k-1} 2^{i(d/2-p)}.$$

Focusing once again on the "error terms", we use the inequality

$$\sum_{i=0}^{k-1} 2^{i(d/2-p)} = \frac{2^{k(d/2-p)} - 1}{2^{d/2-p} - 1} \leq \frac{1}{2^{d/2-p} - 1} m^{d/2-p}.$$

to deduce that

$$\sqrt{2d^p}\lambda^{p-d/2}\sum_{i=0}^{k-1} 2^{i(d/2-p)} \leq \frac{\sqrt{2d^p}}{2^{d/2-p} - 1}(\lambda/m)^{p-d/2}.$$

To conclude, we use the fact that, since $i = k$, a cube $Q_{k,j}$ is either contained in $[0,1]^d$ or disjoint from it (hence giving no contribution to the sum). By (5.5), since the cubes that give contribution are m^d, we have

$$\sum_{j=1}^{2^{kd}} \lambda^{p-d}\mathbb{E}\left[b_p(X,Y;Q_{k,j}\right] = f(\lambda/m).$$

Putting all together, we have the inequality, equivalent to (5.3),

$$f(\lambda) \leq f(\lambda/m) + c(\lambda/m)^{d/2-p},$$

with $c = \sqrt{2d^p}/(2^{d/2-p} - 1)$ independent of m or λ. ∎

Remark 5.7. There is a final step that for brevity we do not perform (see [BdMM02, BB13] for a proof) deduce from Theorem 5.6 that

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[B_{p,n}\right]}{n^{1-p/d}} \in (0,\infty),$$

and in fact coincides with the limit for the Poisson case.

6 A PDE approach

In this last section we describe a recent approach [CLPS14, AST19] which (at present) allows one to rigorously deduce rather precise estimates when $d = 2$. In particular, for $d = 2$, $p = 2$, one can prove that

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[B_{2,n}\right]}{\log(n)} = \frac{1}{2\pi}.$$

Notice that the rate $\log(n)$ differs from the heuristic $n^{1-p/d} = n^0$. This phenomenon was already known [AKT84], but the existence of a limit was not. Infact, the following problem is open:

$$\text{does} \quad \lim_{n\to\infty} \frac{\mathbb{E}\left[B_{1,n}\right]}{\sqrt{n\log(n)}} \quad \text{exist?}$$

To give an illustration of the method, in this section we sketch how one can prove that

$$0 < \liminf_{n\to\infty} \frac{\mathbb{E}\left[B_{1,n}\right]}{\sqrt{n\log(n)}} \quad \text{and} \quad \limsup_{n\to\infty} \frac{\mathbb{E}\left[B_{1,n}\right]}{\sqrt{n\log(n)}} < \infty.$$

Notice that, also in this case the heuristic rate should be $n^{1-1/2} = \sqrt{n}$.

The main idea is to consider the bipartite matching problem as a special instance of *optimal transport* problems, which in recent years attracted much interest both from theoretical [Vil09, AGS08] and applied research [San15, PC18]. In particular, it allows us to consider generalized matching problems both for discrete and "continuous" distributions of points. Instead of looking for a permutation σ, one should instead search for a more general way to match points, allowing e.g. multiple matchings: the main idea (due to L. Kantorovich) is to replace permutations $\sigma \in \mathcal{S}^n$ with transition probabilities (Markov kernels), so that one obtains a probabilistic pairing between points. The strength of the approach is that Markov kernels make sense both in discrete and continuous setting, so that the theory can be generalized.

For our purpose, we do not enter in a detailed description, for we will only use a basic (but crucial) *duality* formula, stating that

$$B_{1,n} = \max_{f\in \mathrm{Lip}(1)} \left\{ \sum_{i=1}^n f(X_i) - f(Y_i) \right\}, \tag{6.1}$$

where $\mathrm{Lip}(1)$ denotes the set of $f : [0,1]^d \to \mathbb{R}$ such that $|f(x) - f(y)| \le |x-y|$ for every x, $y \in [0,1]^d$.

Remark 6.1. For general p duality reads

$$B_{p,n} = \max \left\{ \sum_{i=1}^n f(X_i) - g(Y_i) \,:\, f(x) - g(y) \le |x-y|^p \right\}, \tag{6.2}$$

which allows for a nice economical interpretation. Imagine that blue points $(Y_i)_{i=1}^n$ are suppliers of some product and red points $(X_i)_{i=1}^n$ are potential buyers, and because of some (geographical) reason, to transport a unit of good from y to x costs $|x-y|^p$. A transport company offers us to take care of distributing such products. Because of the way they are organized, they are willing to buy a unit of product located at position

y at the price $g(y)$ and sell a unit at position x at price $f(x)$. To be competitive, in general, inequality $f(x) - g(y) \leq |x - y|^p$ must hold (the selling price at x should be smaller than buying the goods at y and transporting from y to x). The term

$$\sum_{i=1}^{n} f(X_i) - g(Y_i)$$

becomes then the overall profit for this company. Formula (6.2) then states that the maximum profit obtainable equals the minimum matching cost.

Back to the random bipartite matching problem, the main idea is to solve the elliptic partial differential equation (PDE)

$$-\Delta u = \sum_{i=1}^{n} (\delta_{X_i} - \delta_{Y_i}) \quad \text{in } [0,1]^2, \tag{6.3}$$

and use it in (6.1). Indeed, we notice that

$$\begin{aligned} \sum_{i=1}^{n} f(X_i) - f(Y_i) &= \int_{[0,1]^2} f \mathsf{d} \left(\sum_{i=1}^{n} (\delta_{X_i} - \delta_{Y_i}) \right) \\ &= - \int_{[0,1]^2} f(x) \Delta u(x) \mathsf{d}x \\ &= \int_{[0,1]^2} \nabla f(x) \nabla u(x) \mathsf{d}x \leq \int_{[0,1]^2} |\nabla u(x)| \mathsf{d}x \end{aligned} \tag{6.4}$$

where in the last line, in the integration by parts we did not consider the boundary terms, and then e.g. assumed that f is continuously differentiable with $|\nabla f(x)| \leq 1$ on $[0,1]^2$. A part from these issues, we exploit the fact that the last expression does not depend on f, so

$$B_{1,n} = \max_{f \in \mathrm{Lip}(1)} \left\{ \sum_{i=1}^{n} f(X_i) - f(Y_i) \right\} \leq \int_{[0,1]^2} |\nabla u(x)| \mathsf{d}x.$$

Taking expectation, we obtain the inequality

$$\mathbb{E}\left[B_{1,n}\right] \leq \mathbb{E}\left[\int_{[0,1]^2} |\nabla u(x)| \mathsf{d}x \right]. \tag{6.5}$$

To be more rigorous, we need at least to specify boundary conditions for u. We choose Neumann boundary conditions

$$\frac{\partial u}{\partial \eta} = 0, \quad \text{on } \partial[0,1]^2,$$

where η denotes the outer normal at $[0,1]^2$, i.e. $\eta(x) = (1,0)$ on $\{0\} \times (0,1)$, $\{1\} \times (0,1)$ and $\eta(x) = (0,1)$ on $(0,1) \times \{0\}$, $(0,1) \times \{1\}$. This way, there are no boundary terms when we integrate by parts in (6.4)

We can solve (6.3) explicitly via Fourier series: a general formula for the solution to $-\Delta u = \mu - \nu$, where μ, ν are measures (with $\mu([0,1]^2) = \nu([0,1]^2)$) is

$$u(x_1, x_2) = 4 \sum_{m \in \mathbb{N}^2} a_m \cos(m_1 \pi x_1) \cos(m_2 \pi x_2),$$

where $m = (m_1, m_2)$, $|m| = \left(m_1^2 + m_2^2\right)^{1/2}$, $a_{(0,0)} = 0$ and

$$a_m = \frac{4}{\pi^2 |m|^2} \int_{[0,1]^2} \cos(m_1 \pi x_1) \cos(m_2 \pi x_2) \mathsf{d}(\mu - \nu)(x_1, x_2).$$

When $\mu = \sum_{i=1}^{n} \delta_{X_i}$, $\nu = \sum_{i=1}^{n} \delta_{Y_i}$ are random, it follows that a_m is a random variables. We can compute mean and variance easily using the following result.

Lemma 6.2. *Let $n \geq 1$, $g : [0,1]^2 \to \mathbb{R}$ be continuous. Then the random variable*

$$\int_{[0,1]^2} g \mathsf{d}\left(\sum_{i=1}^n \delta_{X_i} - \delta_{Y_i}\right) = \sum_{i=1}^n \left(g(X_i) - g(Y_i)\right)$$

has zero mean and variance

$$2n\left(\int_{[0,1]^2} g^2(x)\mathsf{d}x - \left(\int_{[0,1]^2} g(x)\mathsf{d}x\right)^2\right).$$

Proof. Clearly, the mean is zero, for $\mathbb{E}\left[g(X_i)\right] = \mathbb{E}\left[g(Y_i)\right] = \int_{[0,1]^2} g(x)\mathsf{d}x$. For the variance, notice that it is a sum of $2n$ independent random variables, each with variance (the minus sign in $-g(Y_i)$ does not matter)

$$\mathbb{E}\left[g^2(X_i)\right] - \mathbb{E}\left[g(X_i)\right]^2 = \int_{[0,1]^2} g^2(x)\mathsf{d}x - \left(\int_{[0,1]^2} g(x)\mathsf{d}x\right)^2.$$

□

As a consequence, since $g(x_1, x_2) = 4\cos(m_1\pi x_1)\cos(m_2\pi x_2)$ is such that

$$\int_{[0,1]^2} g(x)\mathsf{d}x = 0, \quad \int_{[0,1]^2} g^2(x)\mathsf{d}x = 1,$$

we have that (for $m \neq (0,0)$),

$$\mathbb{E}\left[a_m^2\right] = \frac{2n}{\left(\pi^2 |m|^2\right)^2}.$$

By Cauchy-Schwarz inequality and Plancherel formula

$$\begin{aligned}
\int_{[0,1]^2} |\nabla u(x)|\, \mathsf{d}x &\leq \left(\int_{[0,1]^2} |\nabla u(x)|^2 \mathsf{d}x\right)^{1/2} \\
&= \left(\int_{[0,1]^2} \left(\frac{\partial u}{\partial x_1}(x_1,x_2)\right)^2 + \left(\frac{\partial u}{\partial x_1} u(x_1,x_2)\right)^2 \mathsf{d}x\right)^{1/2} \\
&= \left(\sum_{m\in\mathbb{N}^2} \pi^2 |m|^2 a_m^2\right)^{1/2}.
\end{aligned}$$

Taking expectation and using Cauchy-Schwarz inequality once more,

$$\begin{aligned}
\mathbb{E}\left[\int_{[0,1]^2} |\nabla u(x)|\, \mathsf{d}x\right] &\leq \mathbb{E}\left[\left(\sum_{m\in\mathbb{N}^2} \pi^2 |m|^2 a_m^2\right)^{1/2}\right] \\
&\leq \mathbb{E}\left[\sum_{m\in\mathbb{N}^2} \pi^2 |m|^2 a_m^2\right]^{1/2} \\
&= \left(\sum_{m\in\mathbb{N}^2\setminus(0,0)} \frac{2n}{\pi^2 |m|^2}\right)^{1/2}.
\end{aligned}$$

Unfortunately, the series we found is divergent, as one can see by comparison with the integral

$$\sum_{m\in\mathbb{N}^2\setminus(0,0)} \frac{2n}{\pi^2 |m|^2} \approx \int_1^\infty \int_0^{\pi/2} \frac{2n}{\pi^2 r^2} r \mathsf{d}r \mathsf{d}\theta = \frac{n}{\pi}\int_1^\infty \frac{1}{r}\mathsf{d}r = +\infty.$$

To overcome this issue we need to carefully approximate u with a smoothed version obtained by “filtering” the high frequencies out, e.g. by restricting summation to $m \in \mathbb{N}^2$ with $|m|$ smaller than a function $h(n) \to \infty$ to be suitably chosen:

$$u_h(x_1, x_2) := 4 \sum_{\substack{m \in \mathbb{N}^2 \\ |m| \le h(n)}} a_{(m_1,m_2)} \cos(m_1 \pi x_1) \cos(m_2 \pi x_2).$$

This of course produces some error terms in (6.5). Once we estimate them, it turns out that one can choose approximatively $h(n) = n^{1/2}$ (which corresponds to smoothing with respect to the space variable on a radius of order $n^{-1/2}$). This way, the partial sums still diverges, but in a controlled way:

$$\sum_{\substack{m \in \mathbb{N}^2 \setminus (0,0) \\ |m| \le n^{1/2}}} \frac{2n}{\pi^2 |m|^2} \approx \int_1^{n^{1/2}} \int_0^{\pi/2} \frac{2n}{\pi^2 r^2} r \mathsf{d}r \mathsf{d}\theta = \frac{n}{\pi} \int_1^{n^{1/2}} \frac{1}{r} \mathsf{d}r = \frac{n \log(n)}{2\pi}.$$

This ultimately leads to the upper bound

$$\limsup_{n \to \infty} \frac{\mathbb{E}[B_{1,n}]}{\sqrt{n \log(n)}} \le \frac{1}{\sqrt{2\pi}}.$$

Concerning the lower bound, going back to the duality formula (6.2) and the formal derivation (6.4), one is tempted to choose

$$f(x) = \frac{u(x)}{\sup_{y \in [0,1]^2} |\nabla u(y)|}.$$

This way, however, it would be difficult to estimate the expectation of the resulting random variable:

$$\mathbb{E}\left[\frac{\int_{[0,1]^2} |\nabla u(x)|^2 \mathsf{d}x}{\sup_{y \in [0,1]^2} |\nabla u(y)|} \right] \le \mathbb{E}[B_{1,n}].$$

Moreover, we already know that we should use u_h instead of u (with $h(n) = n^{1/2}$). Clearly, u_h is smooth, but it turns out that, with some effort, one can estimate from above

$$\mathbb{E}\left[\int_{[0,1]^2} |\nabla u_h(x)|^4 \mathsf{d}x \right]^{1/4} \le c\sqrt{n \log(n)}.$$

where $c \ge 0$ is some constant (independent of $n \ge 1$).

Next, we exploit a classical result from the theory of Sobolev functions [Liu77].

Theorem 6.3. *For every $p > 1$ there exists $c_p > 0$ such that the following holds. For every $g : [0,1]^2 \to \mathbb{R}$ be C^1, there exists $f : [0,1]^2 \to \mathbb{R}$ such that*

$$f \in \mathrm{Lip}(1) \quad \textit{and} \quad \left| \left\{ x \in [0,1]^2 \, : \, f(x) \neq g(x) \right\} \right| \le c_p \int_{[0,1]^2} |\nabla g|^p (x) \mathsf{d}x.$$

To obtain the lower bound, we choose $M > 0$ (to be specified later) and apply the result with $p = 4$ and $g = u_h / (M\sqrt{n \log(n)})$, obtaining $f \in \mathrm{Lip}(1)$ so that (after taking care of the substitution of u with u_h),

$$\begin{aligned}
B_{1,n} \ge \sum_{i=1}^n f(X_i) - f(Y_i) &= -\int_{[0,1]^2} f(x) \Delta u_h(x) \mathsf{d}x \\
&= \int_{[0,1]^2} \nabla f(x) \nabla u_h(x) \mathsf{d}x \\
&= \int_{\{f=g\}} \nabla g(x) \nabla u_h(x) \mathsf{d}x + \int_{\{f \neq g\}} \nabla f(x) \nabla u_h(x) \mathsf{d}x \\
&= \int_{[0,1]^2} \nabla g(x) \nabla u_h(x) \mathsf{d}x - \int_{\{f \neq g\}} \nabla g(x) \nabla u_h(x) \mathsf{d}x + \int_{\{f \neq g\}} \nabla f(x) \nabla u_h(x) \mathsf{d}x
\end{aligned}$$

The first term in the sum above equals

$$\frac{M}{\sqrt{n\log(n)}}\int_{[0,1]^2}|\nabla u_h(x)|^2 \mathsf{d}x,$$

so that its expectation divided by $\sqrt{n\log(n)}$ converges:

$$\lim_{n\to\infty}\frac{1}{Mn\log(n)}\int_{[0,1]^2}|\nabla u_h(x)|^2 \mathsf{d}x = \frac{1}{M2\pi}.$$

We estimate the remaining terms using Hölder inequality

$$\begin{aligned}\left|\int_{\{f\neq g\}}\nabla g(x)\nabla u_h(x)\mathsf{d}x\right| &\leq \left(|f\neq g|^2\int_{[0,1]^2}|\nabla g(x)|^4\mathsf{d}x\int_{[0,1]^2}|\nabla u_h(x)|^4\mathsf{d}x\right)^{1/4}\\ &\leq \frac{c^{1/2}}{M^{-3}(n\log(n))^{3/2}}\int_{[0,1]^2}|\nabla u_h(x)|^4\mathsf{d}x,\end{aligned}$$

so that in expectation

$$\mathbb{E}\left[\left|\int_{\{f\neq g\}}\nabla g(x)\nabla u_h(x)\mathsf{d}x\right|\right]\leq\frac{c\sqrt{n\log(n)}}{M^3},$$

where $c \geq 0$ denotes a different constant (independent of n and M). One obtains a similar estimate for the third term. It follows that

$$\liminf_{n\to\infty}\frac{\mathbb{E}\left[B_{1,n}\right]}{\sqrt{n\log(n)}}\geq\frac{1}{M2\pi}-\frac{1}{M^3}>0$$

by choosing M large enough.

References

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR 2401600

[AKT84] M. Ajtai, J. Komlós, and G. Tusnády, *On optimal matchings*, Combinatorica **4** (1984), no. 4, 259–264. MR 779885

[AST19] Luigi Ambrosio, Federico Stra, and Dario Trevisan, *A PDE approach to a 2-dimensional matching problem*, Probab. Theory Related Fields **173** (2019), no. 1-2, 433–477. MR 3916111

[BB13] Franck Barthe and Charles Bordenave, *Combinatorial optimization over two random point sets*, Séminaire de Probabilités XLV, Lecture Notes in Math., vol. 2078, Springer, Cham, 2013, pp. 483–535. MR 3185927

[BdMM02] J. H. Boutet de Monvel and O. C. Martin, *Almost sure convergence of the minimum bipartite matching functional in Euclidean space*, Combinatorica **22** (2002), no. 4, 523–530. MR 1956991

[CLPS14] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, *Scaling hypothesis for the Euclidean bipartite matching problem*, Phys. Rev. E **90** (2014), no. 1, 012118.

[Liu77] Fon-Che Liu, *A luzin type property of sobolev functions*, Indiana Univ. Math. J. **26** (1977), 645–651.

[PC18] Gabriel Peyré and Marco Cuturi, *Computational Optimal Transport*, arXiv:1803.00567 [stat] (2018), arXiv: 1803.00567.

[RS13] Maxim Raginsky and Igal Sason, *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*, CIT **10** (2013), no. 1-2, 1–246 (English).

[San15] Filippo Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, Birkhäuser, October 2015 (en), Google-Books-ID: UOHHCgAAQBAJ.

[STAD13] Alonso Silva, Hamidou Tembine, Eitan Altman, and Mérouane Debbah, *Optimum and Equilibrium in Assignment Problems With Congestion: Mobile Terminals Association to Base Stations*, IEEE Transactions on Automatic Control **58** (2013), no. 8, 2018–2031.

[Ste97] J. Michael Steele, *Probability theory and combinatorial optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 69, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. MR 1422018

[Tal96] Michel Talagrand, *A new look at independence*, Ann. Probab. **24** (1996), no. 1, 1–34. MR 1387624

[Vil09] Cédric Villani, *Optimal Transport*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.