Random Euclidean matching problems in one dimension

Sergio Caracciolo[*] and Matteo D'Achille[†]

Dipartimento di Fisica, University of Milan and INFN, via Celoria 16, I-20133 Milan, Italy

Gabriele Sicuro[‡]

Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, I-00185, Rome, Italy

(Dated: July 19, 2017)

We discuss the optimal matching solution for both the assignment problem and the matching problem in one dimension for a large class of convex cost functions. We consider the problem in a compact set with the topology both of the interval and of the circumference. Afterwards, we assume the points' positions to be random variables identically and independently distributed on the considered domain. We analytically obtain the average optimal cost in the asymptotic regime of very large number of points N and some correlation functions for a power-law type cost function in the form $c(z) = z^p$, both in the $p > 1$ case and in the $p < 0$ case. The scaling of the optimal cost with the number of points is $N^{p/2}$ for the assignment and N^p for the matching when $p > 1$, whereas in both cases it is a constant when $p < 0$. Finally, our predictions are compared with the results of numerical simulations.

I. INTRODUCTION

Given a generic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ set of edges, a *matching* $\mu \subseteq \mathcal{E}$ on $\mathcal{G}$ is a subset of edges of $\mathcal{G}$ such that, given two edges in μ, they do not share a common vertex [1]. A matching μ is said to be *maximal* if, for any $e \in \mathcal{E} \setminus \mu$, $\mu \cup \{e\}$ is no longer a matching. Denoting by $|\mu|$ the cardinality of μ, we define $\nu(\mathcal{G}) := \max_\mu |\mu|$ the *matching number* of $\mathcal{G}$, and we say that μ is *maximum* if $|\mu| = \nu(\mathcal{G})$. A perfect matching (or 1-factor) is a matching that matches all vertices of the graph. Every perfect matching is also maximum and hence maximal. A perfect matching is also a minimum-size edge cover. We will denote by $\mathcal{M}$ the set of perfect matchings.

Let us suppose now that a weight $w_e \geq 0$ is assigned to each edge $e \in \mathcal{E}$ of the graph $\mathcal{G}$. We can associate to each perfect matching μ a total cost

$$E(\mu) := \sum_{e \in \mu} w_e, \tag{1}$$

and a mean cost per edge

$$\varepsilon(\mu) := \frac{1}{\nu(\mathcal{G})} \sum_{e \in \mu} w_e. \tag{2}$$

In the (weighted) matching problem we search for the perfect matching μ such that the total cost in Eq. (1) is minimized, i.e., the *optimal* matching μ^o is such that

$$E(\mu^o) = \min_{\mu \in \mathcal{M}} E(\mu). \tag{3}$$

In *Euclidean* matching problems, the graph $\mathcal{G}$ is supposed to be embedded in a d-dimensional Euclidean domain $\Lambda \subseteq \mathbb{R}^d$ through an embedding function Φ, in such

a way that each vertex $v \in \mathcal{V}$ of the graph is associated to an Euclidean point $v \mapsto \Phi(v) \in \Lambda$. In this case, the cost w_e associated to the edge $e = (u, v)$ is typically a function of the distance of the images of its corresponding endpoints in Λ, i.e., $w_e = c[\|\Phi(u) - \Phi(v)\|]$. In the following we will refer to the function c as the cost function of the problem. Once the weights are assigned, the problem can be solved using efficient algorithms available in the literature [2, 3].

In a particularly well studied class of matching problems, namely *random* matching problems, the costs $\{w_e\}_e$ are random quantities. In this case, the typical properties of the optimal solution are of a certain interest, and in particular the average optimal cost

$$\overline{E} := \overline{\min_{\mu \in \mathcal{M}} E(\mu)}, \tag{4}$$

where we have denoted by $\overline{\bullet}$ the average over all possible instances of the costs set. The simplest way to introduce randomness in the problem is to consider the weights $\{w_e\}_e$ independent and identically distributed random variables [4, 5]. In random Euclidean matching problems, instead, the randomness is typically introduced associating the vertices of $\mathcal{G}$ to points in Λ that are randomly generated with a given probability distribution density on Λ [6, 7]. Random Euclidean matching problems are usually more difficult to investigate, due to the presence of Euclidean correlations among the weights. A large literature on random optimization problems is nowadays available [3]. After the seminal works of Kirkpatrick *et al.* [8], Orland [9] and Mézard and Parisi [4], random optimization problems have been successfully studied using statistical physics techniques. For example, given a certain instance of a matching problem, we can introduce an "inverse temperature" β and write a partition function

$$Z(\beta) = \sum_{\mu \in \mathcal{M}} e^{-\beta E(\mu)}, \tag{5}$$

in such a way that the optimal cost is recovered as

$$E(\mu^o) = \lim_{\beta \to +\infty} -\frac{\partial \ln Z(\beta)}{\partial \beta}. \tag{6}$$

In the presence of randomness, we have

$$\overline{E} = \lim_{\beta \to +\infty} -\frac{\partial \overline{\ln Z(\beta)}}{\partial \beta}. \tag{7}$$

The treatment of the average appearing in the previous equation has been tackled using the celebrated *replica trick*, that allowed the derivation of fundamental results for the purely random case about both the solution in the thermodynamical limit [4] and the finite-size corrections [10]. The replica approach proved to be successful also in the investigation of random Euclidean matching problems [6, 10]. As an alternative to the classical methods, in Ref. [11] a field theoretical approach has been proposed for the study of the so called quadratic random Euclidean matching problem in any dimension, that was based on a scaling ansatz proposed in Ref. [12] and on the deep connections with optimal transportation theory [13–15].

In the present paper, we will consider the random Euclidean matching problem in one dimension, for a cost function $c(z) = z^p$, with $p \in \mathbb{R} \setminus [0, 1]$, in which the points are independently and uniformly generated on the compact interval $\Lambda = [0, 1]$ and on the unit circumference (some results in the case of noncompact supports and general properties of the convergence rate in the context of optimal transportation theory can be found in the review of Bobkov and Ledoux [15]). We will distinguish between the random Euclidean matching problem and the random Euclidean assignment problem, depending if the embedded graph is the complete graph or the bipartite complete graph respectively.

The Euclidean assignment problem in one dimension has already been extensively studied in the case in which the cost function is convex and increasing [11, 12, 16–19], that is in the case $p > 1$. Here, in this particular case, we will consider also finite size corrections to the average optimal cost that had not been computed previously. More interestingly, we will introduce a general class of cost functions, called here $\mathscr{C}$-functions, that determine a specific structure for the optimal matching solution for a given instance. More precisely, once the two sets of points are labeled in increasing order according to their position along the line, the optimal matching has to be a periodic shift. This is the case of our cost functions also for $p < 0$. So we can derive the average properties of the optimal solution on the basis of this result. Also in this case, as already known for $p > 1$, the optimal solution can be related, in the limiting case of an infinite number of points, to a Gaussian stochastic process.

We also extend the results we obtained for the Euclidean assignment problem to the Euclidean matching problem.

The paper is therefore subdivided in such a way to present in Sec. II the case of the assignment and in Sec. III the case of the matching. Finally, we will give our conclusions in Sec. IV.

II. THE RANDOM EUCLIDEAN ASSIGNMENT PROBLEM

With reference to the definitions given in the Introduction, in the *assignment problem*, we assume $\mathscr{G} = \mathscr{K}_{N,N}$, the complete bipartite graph in which $\mathscr{V} = \mathscr{V}_1 \cup \mathscr{V}_2$, $\mathscr{V}_1 \cap \mathscr{V}_2 = \emptyset$, $|\mathscr{V}_1| = |\mathscr{V}_2| = N$. In the *random Euclidean assignment problem in one dimension*, we consider two sets of points $\Xi_N := \{x_i\}_{i=1,\dots,N}$ and $\Upsilon_N := \{y_j\}_{j=1,\dots,N}$, independently generated with uniform distribution density on $\Lambda = [0, 1]$; we associate then the points in Ξ_N, respectively Υ_N, to the vertices in $\mathscr{V}_1$, respectively $\mathscr{V}_2$. We will assume that the points are labeled in such a way that

$$0 \le x_1 < x_2 < \cdots < x_N \le 1, \tag{8a}$$
$$0 \le y_1 < y_2 < \cdots < y_N \le 1. \tag{8b}$$

A maximum matching $\mu \subset \mathscr{V}_1 \times \mathscr{V}_2$ uniquely corresponds to a permutation $\pi \in \mathcal{S}_N$ of the N elements $[N] := \{1, \dots, N\}$, in such a way that, if $(i, j) \in \mu$, $j = \pi(i)$. We associate to π a *matching cost* and a *mean cost per edge* respectively given by

$$E_N(\pi) := \sum_{i=1}^N w(x_i, y_{\pi(i)}), \tag{9a}$$

$$\varepsilon_N(\pi) := \frac{1}{N} E_N(\pi). \tag{9b}$$

In the previous expressions, the cost function $w(x_i, y_j)$ depends on the points' positions x_i and y_j. As anticipated, we will restrict ourselves to cost functions in the form

$$w(x_i, y_j) := c\left(|x_i - y_j|\right), \quad c \colon \Lambda \to \mathbb{R}^+. \tag{10}$$

We are interested in the asymptotic behavior for $N \gg 1$ of the *average* optimal (mean) cost

$$\varepsilon_N := \overline{\min_\pi \varepsilon_N(\pi)}, \tag{11}$$

where we have denoted by $\overline{\bullet}$ the average over the points' positions. We will show that the typical properties of the solution strongly depend on the properties of the cost function $c(z)$. To study this dependency, we will assume in particular

$$c(z) := z^p, \quad p \in \mathbb{R}, \quad z \in \Lambda. \tag{12}$$

As we will show below, the properties of the optimal solution will depend on the chosen value of p.

A. On the structure of optimal matching

We shall here discuss some general features of the optimal solution in a given instance at variance with the cost function.

1. Preliminaries

Let us first introduce some preliminary definitions and results.

Definition II.1. Given a set of n elements, we say that a permutation $\pi \in \mathcal{S}_n$ of n elements belongs to $\mathcal{C}_n \subseteq \mathcal{S}_n$ if an integer number k exists such that $0 \le k < n$ and

$$\pi(i) = i + k \pmod{n}, \quad i = 1, \ldots, n. \tag{13}$$

Observe that, for $k \ne 0$, a permutation $\pi \in \mathcal{C}_n$ is a cyclic permutation having one cycle only. For $k = 0$ we have the identity permutation, which has n cycles. The set $\mathcal{C}_n \subseteq \mathcal{S}_n$ is an Abelian subgroup corresponding to the cyclic group of the n proper rotations in the plane which leave a regular polygon with n vertices invariant. For $n = 2$, $\mathcal{C}_2 = \{(1)(2), (1,2)\} = \mathcal{S}_2$. The group $\mathcal{C}_3$ coincides with the alternating group of even permutations $\mathcal{A}_3$.

Definition II.2. Given a triple of three integers $(i, j, k) \in [n]^3$, $[n] := \{1, \ldots, n\}$, we say that the ordered triple of integers (a, b, c) is *cyclically co-oriented* with it when an even permutation $\pi \in \mathcal{C}_3$ exists such that $(\pi(a), \pi(b), \pi(c))$ is in the same order of (i, j, k) respect to the order relation of the integers.

The following Proposition, that appeared in Ref. [17], will be fundamental in the following.

Proposition II.1. *A permutation $\pi \in \mathcal{S}_n$ belongs to $\mathcal{C}_n$ if, and only if, for any triple (i, j, k), the corresponding triple $(\pi(i), \pi(j), \pi(k))$ is cyclically co-oriented with (i, j, k).*

Proof. If $\pi \in \mathcal{C}_n$ all the triples $(\pi(i), \pi(j), \pi(k))$ are cyclically co-oriented with (i, j, k), due to the fact that all permutations in $\mathcal{C}_n$ are ordering preserving.

For the converse, let us assume that all triples are cyclically co-oriented with their image through the permutation π. Observe now that if for any couple $(i, j) \in [n]^2$ we have $\pi(i) - \pi(j) = i - j \pmod{n}$, then $\pi \in \mathcal{C}_n$. To prove this statement, we proceed by contradiction and we assume that there exists at least a couple $(i, j) \in [n]^2$ such that $\pi(i) - \pi(j) \ne i - j \pmod{n}$. It follows that the sequences

$$I = (i, i + 1 \pmod{n}, \ldots, j), \tag{14a}$$
$$J = (\pi(i), \pi(i) + 1 \pmod{n}, \ldots, \pi(j)) \tag{14b}$$

have not the same cardinality, and therefore there must exist a k such that either $k \in I$ and $\pi(k) \notin J$ or $k \notin I$ and $\pi(k) \in J$. By consequence, the triples (i, j, k) and $(\pi(i), \pi(j), \pi(k))$ are not cyclically co-oriented, that is in contradiction with the hypothesis and therefore the theorem is proved. $\square$

2. $\mathcal{C}$-functions

As anticipated, an assignment between two sets of N points can be uniquely associated to a permutation of N elements $\pi \in \mathcal{S}_N$. We will show below that the optimal permutations belongs to $\mathcal{C}_N \subseteq \mathcal{S}_N$ if the cost function $c(z)$ appearing in Eq. (10) satisfies the following property.

Definition II.3. We shall say that a function $f : (0,1) \to \mathbb{R}^+$ is a $\mathcal{C}$-*function* if for any $\eta \in [0, 1 - z_2]$, $\hat{\eta} \in [z_2, 1]$ and $0 < z_1 < z_2 < 1$

$$f(z_2) - f(z_1) \le f(\eta + z_2) - f(\eta + z_1), \tag{15a}$$
$$f(z_2) - f(z_1) \le f(\hat{\eta} - z_2) - f(\hat{\eta} - z_1). \tag{15b}$$

Eq. (15a) implies that $\Psi_\eta(z) = f(\eta + z) - f(z)$ is an increasing function in the interval $(0, 1 - \eta)$ for any value of $\eta \in (0, 1)$. Moreover, if f is continuous, Eq. (15a) is equivalent to convexity (see Appendix A).

Eq. (15b) implies that the function $\Phi_\eta(z) := f(\eta - z) - f(z)$ is increasing in the interval $(0, \eta)$ for any value of $\eta \in (0, 1)$. If f is differentiable, this fact can be written as

$$f'(\eta - z) + f'(z) \le 0, \quad z \in (0, \eta), \quad \eta \in (0, 1), \tag{16a}$$

that for $\eta \to 1$ becomes

$$f'(1 - z) + f'(z) \le 0. \tag{16b}$$

This implies, for example, that the convex function

$$f_\alpha(x) = (x - \alpha)^2, \quad \alpha \in \mathbb{R}, \tag{17}$$

is a $\mathcal{C}$-function on Λ for $\alpha \ge 1/2$ only.

3. Optimal matching on a segment

Let us now discuss the structure of the optimal assignment on the line. We start with the following Definition to fix our nomenclature.

Definition II.4 (Crossing and planar matching)**.** Let us consider two sets of points $\Xi_N = \{x_i\}_{i=1,\ldots,N}$ and $\Upsilon_N = \{y_i\}_{i=1,\ldots,N}$ on the interval $\Lambda := [0, 1]$ and let us assume that they are labeled in such a way that if $i < j$ then $x_i < x_j$ and $y_i < y_j$. Then a matching between Ξ_N and Υ_N is said to be *planar*, or *non-crossing* if, given the corresponding permutation π and two pairs of matched points $(x_i, y_{\pi(i)})$ and $(x_j, y_{\pi(j)})$, $i < j$, the corresponding intervals are either disjoint, $x_i < y_{\pi(i)} < x_j < y_{\pi(j)}$, or nested, $x_i < x_j < y_{\pi(j)} < y_{\pi(i)}$. The matching is otherwise said to be *crossing.*

From the pictorial point of view, in a planar matching it is always possible to draw semi-arcs in the upper semi-plane joining the couples of matched points which do not intersect, e.g.,
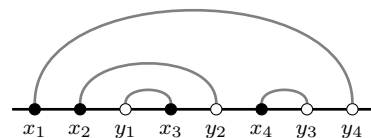
It is well known that the optimal matching configuration is planar if the cost function c in Eq. (10) is concave [16, 20, 21].

In the following, we will restrict ourselves to two classes of convex cost functions, namely $\mathscr{C}$-functions and strictly increasing convex functions. We can start studying in detail the matching problem for $N = 2$, considering two white points and two black points on the line. We can assume, without loss of generality, that the first point along the line is black. There are therefore 3 possible orderings of the points, namely
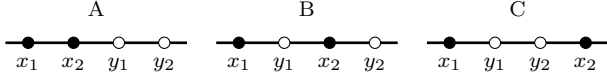
Each configuration allows two possible matchings, namely $\pi(i) = i$ and $\pi(i) = i + 1 \pmod 2$ for $i = 1, 2$. The following Propositions hold.

Proposition II.2. *Given the assignment problem on the interval $(0, 1)$ with $N = 2$, if the cost function $c(z)\colon (0, 1) \to \mathbb{R}^+$ appearing in Eq. (10) is a $\mathscr{C}$-function, then the optimal matching is the crossing one, whenever a crossing matching is available.*

Proof. The proof of the Proposition is straightforward. In the $N = 2$ case only two of the three possible configurations allow a crossing solution, namely the configuration A and the configuration C. The configuration A allows the two possible matchings
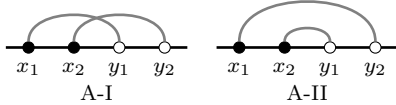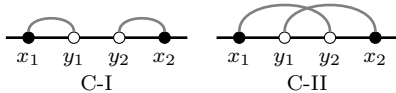

With reference to the picture above,

$$E_{\mathrm{I}}^{\mathrm{A}} = c(y_1 - x_1) + c(y_2 - x_2), \qquad (18a)$$
$$E_{\mathrm{II}}^{\mathrm{A}} = c(y_2 - x_1) + c(y_1 - x_2). \qquad (18b)$$

But, being c a $\mathscr{C}$-function,

$$c(y_1 - x_1) - c(y_1 - x_2) \le c(y_2 - x_1) - c(y_2 - x_2) \quad (19)$$

where we have used Eq. (15a) with $\eta = y_2 - y_1$, and therefore we have that $E_{\mathrm{I}}^{\mathrm{A}} \le E_{\mathrm{II}}^{\mathrm{A}}$. Observe that this result for the A case holds for any continuous convex function $c(z)$ on $(0, 1)$. Similarly, the configuration C allows the two possible matchings

with corresponding costs

$$E_{\mathrm{I}}^{\mathrm{C}} = c(y_1 - x_1) + c(x_2 - y_2), \qquad (20a)$$
$$E_{\mathrm{II}}^{\mathrm{C}} = c(y_2 - x_1) + c(x_2 - y_1). \qquad (20b)$$

Again, being c a $\mathscr{C}$-function,

$$c(y_1 - x_1) - c(y_1 - x_2) \le c(y_2 - x_1) - c(y_2 - x_2) \quad (21)$$

where we have used Eq. (15b) with $\eta = y_2 - x_1$, and therefore $E_{\mathrm{II}}^{\mathrm{C}} \le E_{\mathrm{I}}^{\mathrm{C}}$. This completes the proof. $\square$

Proposition II.3 (Convex increasing function). *Given the assignment problem on the interval Λ, if the cost function $c(z)\colon [0, 1) \to \mathbb{R}^+$ appearing in Eq. (10) is a strictly convex increasing function, then the optimal permutation in the $N = 2$ case is the identity permutation $\pi(i) = i$, $i = 1, 2$.*

Proof. Let us first observe that a strictly increasing function on Λ cannot be a $\mathscr{C}$-function, due to property in Eq. (16). However, we have already proved that, if $c(z)$ is convex, the optimal matching in the configuration A is the ordered one. If we consider now the configuration B, we have to evaluate two possible matchings, namely
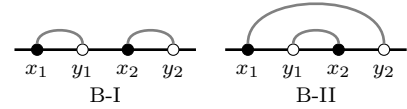
In this case,

$$E_{\mathrm{I}}^{\mathrm{B}} = c(y_1 - x_1) + c(x_2 - y_2), \qquad (22a)$$
$$E_{\mathrm{II}}^{\mathrm{B}} = c(y_2 - x_1) + c(x_2 - y_1). \qquad (22b)$$

By hypothesis, the quantity $c(y_1 - x_1) - c(x_2 - y_1)$ is monotonously increasing respect to the variable $y_1 \in (x_1, x_2)$ and therefore

$$c(y_1 - x_1) - c(x_2 - y_1) \le c(x_2 - x_1) - c(0). \qquad (23)$$

Similarly, convexity implies that $c(y_2 - x_1) - c(y_2 - x_2)$ is increasing in its argument y_2 and therefore

$$c(y_2 - x_1) - c(y_2 - x_2) \ge c(x_2 - x_1) - c(0). \qquad (24)$$

It follows that

$$c(y_1 - x_1) - c(x_2 - y_1) \le c(y_2 - x_1) - c(y_2 - x_2) \Rightarrow$$
$$\Rightarrow E_{\mathrm{I}}^{\mathrm{B}} \le E_{\mathrm{II}}^{\mathrm{B}}. \quad (25)$$

If we finally consider the configuration C, we have

$$E_{\mathrm{I}}^{\mathrm{C}} = c(y_1 - x_1) + c(x_2 - y_2) \le c(y_2 - x_1) + c(x_2 - y_1) = E_{\mathrm{II}}^{\mathrm{C}} \qquad (26)$$

due to the fact that the cost function c is strictly increasing. $\square$

If we consider now the $N = 3$ case, we can derive the following fundamental Lemma.

Lemma II.4. *Let $c(z)\colon (0, 1) \to \mathbb{R}^+$, cost function for the assignment problem on the interval $(0, 1)$, be a $\mathscr{C}$-function. Then the optimal permutation π in the assignment problem for $N = 3$ belongs to the set $\mathscr{C}_3$.*

Proof. For the proof of this Lemma, observe that, by Proposition II.2, for $N = 2$ the crossing matching is the optimal one whenever it is available, i.e., we can obtain the optimal matching maximizing the number of crossings given a certain configuration. In the assignment problem with $N = 3$ "white" points to be matched with $N = 3$ "black" points on the line, there are $\frac{1}{2}\binom{6}{3} = 10$ distinct configurations assuming that the first point is always of a given type, e.g., black. Each configuration allows six possible matching permutations. Let us start with the following three configurations,
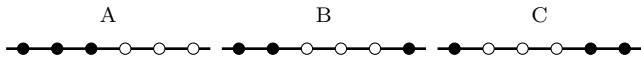

All of them can be pictorially represented, in a compact way, on a circumference as

For example, we can represent
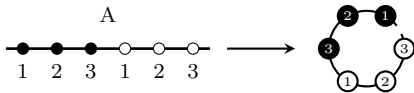
In this representation, each matching on a configuration on the line corresponds to a set of three chords joining the points on the circumference, and a crossing in a matching corresponds to an actual crossing between chords. We will use this representation to evaluate all configurations in a compact way. By Proposition II.2, we can order the possible matchings using the fact that, applying a transposition to a given permutation, if a new crossing appears, then the new permutation has a lower cost. Pictorially we can order the six permutations as
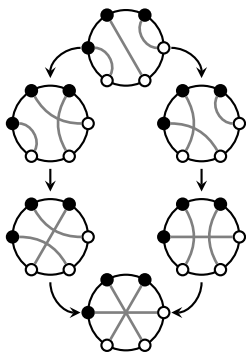
where the arrows denote the transition from a given matching to another one with lower cost, as a consequence of a single transposition. In this case the final, and cheapest, matching corresponds to the optimal permutation $\pi(i) = i$ for the configuration A, $\pi(i) = i + 1 \pmod 3$ for the configuration B, and $\pi(i) = i + 2 \pmod 3$ for the configuration C.
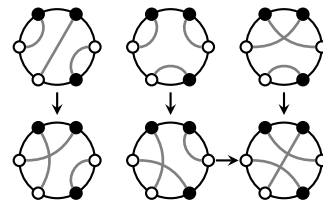
The following six configurations of points
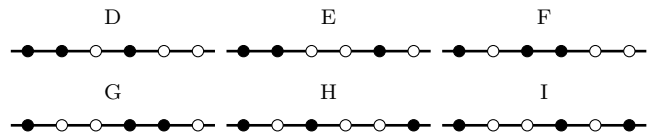
can be treated similarly. In particular, all six configurations can be represented as
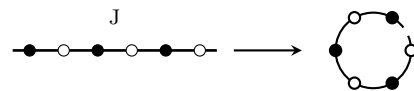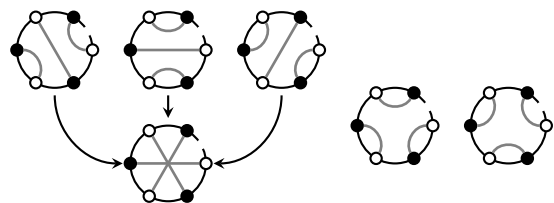

We have therefore

This implies that, for each configuration, there are two possible optimal permutations, namely $\pi(i) = i$ or $\pi(i) = i + 1 \pmod 3$ for the configurations D, E, F, and $\pi(i) = i+1 \pmod 3$ or $\pi(i) = i+2 \pmod 3$ for the configurations G, H, I.

Finally, let us consider the configuration

In this case we have

There are three possible optimal permutations, namely $\pi(i) = i$, $\pi(i) = i + 1 \pmod 3$ and $\pi(i) = i + 2 \pmod 3$.

Collecting our results, we have that, in the $N = 3$ case, the optimal permutation π is such that $\pi \in \mathscr{C}_3$. $\qquad\square$

Proposition II.3 allows us to state the following Theorem, that generalizes an analogous one proven by Boniolo *et al.* [17] in the particular case of the cost function $c(z) = z^p$ with $p > 1$. An equivalent statemant for general convex increasing functions can be found, for example, in Ref. [16].

Theorem II.5 (Optimal matching with convex increasing cost function)**.** *Given the assignment problem on Λ, if the cost function $c(z)\colon [0,1) \to \mathbb{R}^+$ appearing in Eq. (10) is a strictly convex increasing function, then the optimal permutation is the identity permutation.*

Proof. Let us assume by contradiction that the optimal matching π is not the one corresponding to the identity permutation. Therefore, there exists at least a couple of matched pairs $(i, \pi(i))$, $(j, \pi(j))$ such that $i < j$ and $\pi(i) > \pi(j)$. But, by Proposition II.2, this implies that the cost can be decreased considering instead the matched pairs $(i, \pi(j))$ and $(j, \pi(i))$, hence the absurd and therefore the optimal matching is given by the identity permutation $\pi(i) = i$ for all values of $i = 1, \ldots, N$. $\square$

If the cost function is a $\mathscr{C}$-function, the following Theorem holds.

Theorem II.6 (Optimal matching with $\mathscr{C}$-function). *Given the assignment problem on the interval Λ, if the cost function $c(z)\colon [0,1) \to \mathbb{R}^+$ appearing in Eq. (10) is a $\mathscr{C}$-function, then the optimal permutation π is such that $\pi \in \mathscr{C}_N$.*

Proof. Given the optimal permutation π, because of Lemma II.4, in every subset of three matched couples they must be co-oriented. Therefore, because of Proposition II.1, $\pi \in \mathscr{C}_N$. $\square$

B. Average properties of the optimal solution

In the following, we will apply the previous results to the random Euclidean assignment problem with cost function $c(z) = z^p$, $p \in \mathbb{R} \setminus [0,1]$. We will study the average properties of the optimal solution assuming that the points are uniformly and independently distributed on the unit interval. To stress the dependency on p, we will denote the matching cost and the mean cost per edge corresponding to the permutation $\pi \in \mathcal{S}_N$ by

$$E_N^{(p)}(\pi) := \sum_{i=1}^{N} |x_i - y_{\pi(i)}|^p, \tag{27a}$$

$$\varepsilon_N^{(p)}(\pi) := \frac{1}{N} E_N^{(p)}(\pi), \tag{27b}$$

respectively. The average optimal cost will be given by

$$\varepsilon_N^{(p)} := \overline{\min_{\pi \in \mathcal{S}_N} \varepsilon_N^{(p)}(\pi)}. \tag{28}$$

For this particular case, Theorem II.6 allows us to state the following

Corollary II.7. *Given the assignment problem on the interval Λ with cost function $c(z) = z^p$, denoting by π the optimal permutation, then $\pi(i) = i$ for $p > 1$, and $\pi \in \mathscr{C}_N$ for $p < 0$.*

Proof. It is enough to observe that, for $p > 1$, the function $c(z) = z^p$ is a strictly increasing convex function, and apply therefore Theorem II.5. On the other hand, for $p < 0$, the function $c(z)$ is a $\mathscr{C}$-function, and we can apply Theorem II.6. $\square$

If periodic boundary conditions are assumed (i.e., the problem is considered on the unit circumference), we can derive a similar result. Indeed, the assignment problem with cost function $c(z)$ on the circumference can be restated as an assignment problem on Λ with a modified cost function taking into account the periodicity. In particular, the assignment problem on the circumference with cost function $c(z)$ can be thought as an assigment problem on Λ with cost function

$$\hat{c}(z) = c(z)\theta\left(\frac{1}{2} - z\right)\theta(z) + c(1 - z)\theta(1 - z)\theta\left(z - \frac{1}{2}\right), \tag{29}$$

where $\theta(x)$ is the Heaviside step function. In this case, the following Corollary holds.

Corollary II.8. *Let us consider the assignment problem on the circumference with cost function $c(z) = z^p$. Denoting by π the optimal permutation, then $\pi \in \mathscr{C}_N$ for $p < 0$ or $p > 1$. In particular, for $p = 1$ there exists a cyclic optimal solution.*

Proof. As stated above, the assignment problem on the circumference with cost function $c(z)$ corresponds to the assignment problem on Λ with cost function

$$\hat{c}(z) = z^p\theta(z)\theta\left(\frac{1}{2} - z\right) + (1 - z)^p\theta(1 - z)\theta\left(z - \frac{1}{2}\right). \tag{30}$$

The function above is a $\mathscr{C}$-function for $p < 0$. Indeed, it is easily seen that $\Psi_\eta(z) = \hat{c}(\eta + z) - \hat{c}(z)$ is an increasing function on the interval $(0, 1 - \eta)$ for any value of $\eta \in (0,1)$, and therefore Eq. (15a) holds. Moreover, the function $\Phi_\eta(z) = \hat{c}(\eta - z) - \hat{c}(z)$ is monotonically increasing on the interval $(0, \eta)$, and therefore Eq. (15b) is satisfied. The proof for the $p \geq 1$ case has been given in Ref. [17]. $\square$

1. Donsker's theorem and the Brownian bridge process

In Corollary II.7 and Corollary II.8 we have proved that, for $p \in \mathbb{R} \setminus [0,1]$, the optimal permutation has the form $\pi(i) = i + k \pmod{N}$, for some $k \in [N]$ depending on the instance of our problem. In the case of open boundary conditions (i.e., of the problem on the interval), the optimal cost can be written as

$$\min_{\pi \in \mathcal{S}_N} \varepsilon_N^{(p)}(\pi) = \min_{k \in [N]} \frac{1}{N} \sum_{i=1}^{N} \left| x_i - y_{i+k \pmod{N}} \right|^p, \quad (31)$$

In particular, for $p > 1$, the optimal permutation of the assignment problem on Λ is always $\pi(i) = i$, independently from both the instance and the specific values of p, and therefore the optimal cost is simply given by

$$\min_{\pi \in \mathcal{S}_N} \varepsilon_N^{(p)}(\pi) = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^p. \tag{32}$$

These results imply that the optimal solution is related, in the $N \to \infty$ limit, to a linear combination of two Brownian bridge processes, a fact that follows from Donsker's theorem [22].

Theorem II.9 (Donsker). *For any $N \in \mathbb{N}$, there exists a probability space Ω_N such that we can define on it the random variable $\mathbf{X}_N \coloneqq (\mathsf{X}_i)_i$, $\mathbf{X}_N \colon \Omega_N \to \Lambda^N$, each component X_i being a random variable uniformly distributed on the unit interval Λ. Moreover, let us consider the corresponding Nth empirical process*

$$\mathsf{F}_N(t, \mathbf{X}_N) \coloneqq \frac{1}{N} \sum_{i=1}^{N} \theta(t - \mathsf{X}_i) - t. \qquad (33)$$

Then we can find a sample-continuous Brownian bridge process on Λ, defined on the same probability space Ω_N, $\mathsf{B}_N \colon \Lambda \times \Omega_N \to \mathsf{B}_N(t; \omega)$, such that, for all $\epsilon > 0$,

$$\lim_N \Pr\left[\sup_{t \in [0,1]} \left| \sqrt{N} \mathsf{F}_N(t, \mathbf{X}_N(\omega)) - \mathsf{B}_N(t; \omega) \right| > \epsilon \right] = 0. \qquad (34)$$

Donsker's theorem expresses the (weak) convergence of the process F_N to a Brownian bridge process in the $N \to \infty$ limit. The convergence rate has been studied by Komlós *et al.* [23], that proved that

$$\sup_{0 \le t \le 1} \left| \sqrt{N} \mathsf{F}_N(t, \mathbf{X}_N(\omega)) - \mathsf{B}_N(t; \omega) \right| = O\left(\frac{\ln N}{\sqrt{N}} \right) \qquad (35)$$

almost surely [24].

Let $\Xi_N = \{x_i\}_i \equiv \mathbf{X}_N(\omega)$ be a realization of $\mathbf{X}_N$ for a given instance of the problem $\omega \in \Omega_N$. The empirical process F_N is given by

$$\mathsf{F}_N(t, \mathbf{X}_N(\omega)) = \frac{1}{N} \sum_{i=1}^{N} \theta(t - x_i) - t \qquad (36)$$

Supposing now that the elements of Ξ_N are labeled in such a way that $x_i < x_j \Leftrightarrow i < j$, we have

$$\lim_{t \to x_i^+} \mathsf{F}_N(t, \mathbf{X}_N(\omega)) = \frac{i}{N} - x_i. \qquad (37)$$

Given therefore two realizations ω and $\tilde{\omega}$, corresponding to $\Xi_N = \{x_i\}_i$ and $\Upsilon_N = \{y_i\}_i \equiv \mathbf{X}_N(\tilde{\omega})$ respectively, both generated as above, we can write

$$y_j - x_i = \frac{j - i}{N} + \lim_{t \to x_i^+} \mathsf{F}_N(t, \mathbf{X}_N(\omega)) - \lim_{t \to y_j^+} \mathsf{F}_N(t, \mathbf{X}_N(\tilde{\omega})). \qquad (38)$$

Denoting by $i = Nu + \text{\textonehalf}$ and $j = Nv + \text{\textonehalf}$, $u, v \in \Lambda$, and observing that, for large N and $\epsilon > 0$,

$$\Pr\left[|x_i - u| > \epsilon \right] \simeq \exp\left(-\frac{N \epsilon^2}{2u(1 - u)} \right) \qquad (39)$$

Donsker's theorem allows us to write

$$\sqrt{N} \left(y_{Nv + \text{\textonehalf}} - x_{Nu + \text{\textonehalf}} + u - v \right)$$
$$\xrightarrow{N \to +\infty} \mathsf{B}(u; \omega) - \mathsf{B}(v; \tilde{\omega}), \qquad (40)$$

where the limit is intended in probability. This results implies that we can write the arguments of the sums in Eq. (31) and Eq. (32) in terms of Brownian bridge processes in the large N limit.

2. Open boundary conditions

To be more precise, we start analyzing the random Euclidean assignment problem on the interval, assuming a cost function $c(z) = z^p$ with $p \in \mathbb{R} \setminus [0, 1]$.

a. The $p > 1$ case Many aspects of the random Euclidean assignment problem in one dimension for $p > 1$ have been analyzed in Refs. [16–18]. As observed above, the optimal permutation in this case is always the identity one, $\pi(i) = i$. Denoting by

$$\varphi_i \coloneqq y_i - x_i, \qquad (41)$$

we can write, for any instance and any value of N,

$$\min_{\pi \in \mathcal{S}_N} \varepsilon_N^{(p)}(\pi) = \frac{1}{N} \sum_{i=1}^{N} |\varphi_i|^p. \qquad (42)$$

From Donsker's theorem, we know that in the $N \to +\infty$ limit,

$$\phi(s) \coloneqq \sqrt{N} \varphi_{Ns + \text{\textonehalf}} \xrightarrow{N \to +\infty} \mathsf{B}(s; \omega) - \mathsf{B}(s; \tilde{\omega}), \qquad (43)$$

where we have introduced the new variable s such that $i = Ns + \text{\textonehalf}$. The authors of Refs. [17, 18] used this correspondence between the Brownian bridge process and the optimal solution of the Euclidean assignment problem to derive the expression of the average optimal cost and the correlation function in the $N \to +\infty$ limit. They obtained

$$\lim_{N \to \infty} N^{\frac{p}{2}} \varepsilon_N^{(p)} = \frac{\Gamma\left(1 + \frac{p}{2}\right)}{p + 1} + O\left(\frac{1}{N}\right), \qquad (44)$$

whereas the correlation function has been studied in Refs. [17]. Here we derive

$$c_p(r) \coloneqq \frac{1}{\mathcal{N}(r)} \iint_0^1 \overline{\phi(s)\phi(t)} \delta\left(|s - t| - r\right) \mathrm{d}s\,\mathrm{d}t$$
$$= \iint_0^1 \frac{\overline{\phi(s)\phi(t)} - \overline{\phi(s)}\,\overline{\phi(t)}}{\mathcal{N}(r)} \delta\left(|s - t| - r\right) \mathrm{d}s\,\mathrm{d}t$$
$$= \frac{1}{3}(1 - r)^2, \quad r \in \Lambda, \qquad (45)$$

where

$$\mathcal{N}(r) \coloneqq \iint_0^1 \delta\left(|s - t| - r\right) \mathrm{d}s\,\mathrm{d}t = 2(1 - r). \qquad (46)$$

The correlation function has been obtained using the following fundamental property of the Brownian bridge process,

$$\overline{\mathsf{B}(s;\omega)\mathsf{B}(t;\omega)} = \min\{s,t\} - st. \tag{47}$$

In the following, we derive the finite size corrections to the asymptotic cost in Eq. (44), through a straightforward computation on the optimal matching solution and following the approach of Boniolo *et al.* [17]. Let us first observe that the probability of finding x_k in the interval $(x, x + \mathrm{d}\,x)$ is

$$\Pr[x_k \in \mathrm{d}\,x] = B(k;N,x)\frac{k\,\mathrm{d}\,x}{x}, \quad k = 1,\ldots,N, \tag{48}$$

where we have introduced the binomial distribution

$$B(k;N,p) := \binom{N}{k}p^k(1-p)^{N-k}. \tag{49}$$

In Eq. (48) we have used the notation $x_k \in \mathrm{d}\,x \Leftrightarrow x_k \in (x, x + \mathrm{d}\,x)$. Due to the fact that the random variables x_k and y_k are independent, we can write

$$\Pr[\varphi_k \in \mathrm{d}\,\varphi] = \mathrm{d}\,\varphi \binom{N}{k}^2 k^2 \times$$
$$\times \int\limits_0^1\!\!\int \delta(\varphi - y + x)(xy)^{k-1}[(1-x)(1-y)]^{N-k}\,\mathrm{d}\,x\,\mathrm{d}\,y. \tag{50}$$

Eq. (50) allows in principle the calculation of the average optimal cost for any value N. For example, in the $p = 2$ case, we obtain

$$\min_\pi \varepsilon_N^{(2)}(\pi) = \frac{1}{3(1+N)} = \frac{1}{3N} - \frac{1}{3N^2} + o\left(\frac{1}{N^2}\right). \tag{51}$$

The calculations, however, greatly simplify in the $N \to +\infty$ limit. Introducing the variable $\phi(s) := \sqrt{N}\varphi_{Ns+1/2}$, Eq. (50) can be written, up to higher order terms, as

$$\Pr[\phi(s) \in \mathrm{d}\,\phi] =$$
$$= \mathrm{d}\,\phi\,\frac{\mathrm{e}^{-\frac{\phi^2}{4s(1-s)}}}{2\sqrt{\pi s(1-s)}}\left\{1 + \frac{s(1-s)+1}{8Ns(1-s)} + \frac{7s(1-s)-2}{8Ns^2(1-s)^2}\phi^2\right.$$
$$\left. + \frac{1-5s(1-s)}{32Ns^3(1-s)^3}\phi^4 + o\left(\frac{1}{N}\right)\right\}, \tag{52}$$

see Appendix B. As expected, the leading term is the distribution of a Brownian bridge process on the domain Λ [25]. From Eq. (52) we can easily obtain

$$\lim_{N\to\infty} N^{\frac{p}{2}}\varepsilon_N^{(p)} = \int_0^1 \overline{|\phi(s)|^p}\,\mathrm{d}\,s$$
$$= \frac{\Gamma\left(1 + p/2\right)}{p+1}\left(1 - \frac{1}{N}\frac{p(p+2)}{8}\right) + o\left(\frac{1}{N}\right). \tag{53}$$

The expression of the leading term in Eq. (53) has been numerically verified, for example, in Refs. [17, 18]. In Fig. 2a we compare the results of our simulations with the theoretical prediction given in Eq. (53).

b. The $p < 0$ case Let us now consider the $p < 0$ and let us define

$$\frac{1}{\sqrt{N}}\phi_t^{(N)}(s) :=$$
$$= y_{N(s+t \pmod 1)+1/2} - x_{Ns+1/2} - \sigma(s,t), \tag{54}$$

where $Ns + 1/2 = k \in [N]$ and $Nt \in [N]$ and

$$\sigma(s,t) := (s + t \pmod 1) - s, \quad s,t \in \Lambda. \tag{55}$$

Corollary II.7 states that, for a given instance of our problem, the optimal solution corresponds to a certain value t such that

$$\min_\pi \varepsilon_N^{(p)}(\pi) =$$
$$= \frac{1}{N}\sum_{k=1}^N \left|\sigma\left(\frac{k-1/2}{N},t\right) + \frac{1}{\sqrt{N}}\phi_t^{(N)}\left(\frac{k-1/2}{N}\right)\right|^p. \tag{56}$$

From Donsker's theorem we have

$$\phi_t^{(N)}(s) \xrightarrow{N\to+\infty} \phi_t(s) := \mathsf{B}(s;\omega) - \mathsf{B}\left(s + t \pmod 1; \tilde{\omega}\right), \tag{57}$$

and the optimal cost can be written, in the large N limit, as

$$\frac{1}{N}\sum_{k=1}^N \left|\sigma\left(\frac{k-1/2}{N},t\right) + \frac{1}{\sqrt{N}}\phi_t^{(N)}\left(\frac{k-1/2}{N}\right)\right|^p \xrightarrow{N\gg1}$$
$$\int_0^1 \left|\sigma(s,t) + \frac{\phi_t^{(N)}(s)}{\sqrt{N}}\right|^p\,\mathrm{d}\,s, \tag{58}$$

for some value of t depending both on p and on the specific instance of the problem. The value of t for the optimal solution can be found by minimizing the expression above respect to t. We proceed perturbatively, observing that

$$\lim_N \int_0^1 \left|\sigma(s,t) + \frac{\phi_t^{(N)}(s)}{\sqrt{N}}\right|^p\,\mathrm{d}\,s = t^p(1-t) + t(1-t)^p, \tag{59}$$

that is minimized by $t = 1/2$. To evaluate the nontrivial finite size corrections, we assume therefore

$$t := \frac{1}{2} + \frac{\tau}{\sqrt{N}}, \tag{60}$$

where τ depends both on p and on the instance of the problem. Performing a large N expansion, the cost can be written as

$$\int_0^1 \left| \sigma(s,t) + \frac{\phi_t^{(N)}(s)}{\sqrt{N}} \right|^p ds$$

$$= \frac{1}{2^p} + p \int_0^1 \frac{\text{sign}(1/2 - s)}{2^{p-1}\sqrt{N}} \phi_{1/2}^{(N)}(s)\, d s$$

$$- \frac{p\tau\left(\tau + \phi_{1/2}(1/2)\right)}{2^{p-2}N} + \frac{p(p-1)}{2^{p-1}N} \int_0^1 \left(\tau + \phi_{1/2}(s)\right)^2 d s$$

$$+ \tau p \int_0^1 \frac{\text{sign}(1/2 - s)}{2^{p-1}N} \left. \partial_t \phi_t(s)\right|_{t=\frac{1}{2}} d s + o\left(\frac{1}{N}\right). \quad (61)$$

Here and in the following we adopt the convention $\text{sign}(0) = 1$. Observe that, due to Eq. (35), we can neglect the corrections to the asymptotic limit given in Eq. (57) in all terms appearing in the expansion above, except in the second one, that must be treated differently when the average will be performed, due to the different scaling of the coefficient. Being $\phi_t(s) = \mathsf{B}(s;\omega) - \mathsf{B}(s + t \pmod 1; \hat\omega)$, we can formally write in the last term $\partial_t \phi_t(s) = -\partial_t \mathsf{B}(s + t \pmod 1, \hat\omega) = -\partial_s \mathsf{B}(s + t \pmod 1, \hat\omega)$ and therefore, after an integration by parts, the expression above becomes

$$\int_0^1 \left| \sigma(s,t) + \frac{\phi_t^{(N)}(s)}{\sqrt{N}} \right|^p d s =$$

$$= \frac{1}{2^p} - \frac{p\tau\left(\tau + \phi_{1/2}(1/2) + \phi_{1/2}(0)\right)}{2^{p-2}N}$$

$$+ \frac{p(p-1)}{2^{p-1}N} \int_0^1 \left(\tau + \phi_{1/2}(s)\right)^2 d s$$

$$+ \frac{p}{2^{p-1}\sqrt{N}} \int_0^1 \text{sign}(1/2 - s)\, \phi_{1/2}^{(N)}(s)\, d s + o\left(\frac{1}{N}\right). \quad (62)$$

Minimizing respect to τ, we obtain, up to higher terms,

$$\tau = -\frac{p-1}{p-3} \int_0^1 \phi_{1/2}(s)\, d s + \frac{\phi_{1/2}(1/2) + \phi_{1/2}(0)}{p-3}. \quad (63)$$

We have verified our assumption in Eq. (60) and therefore Eq. (63). In particular, Eq. (63) implies $\overline{\tau} = 0$ and [26]

$$\overline{\tau^2} = \frac{p^2 - 5p + 7}{6(p-3)^2}. \quad (64)$$

The results of our numerical simulations, given in the inset in Fig. 2b, show a good agreement between the prediction in Eq. (64) and simulations.

To obtain the average optimal cost, we have to substitute Eq. (63) into Eq. (62), and then average over the possible realizations, using the fact that, as consequence of Eq. (47), the following property holds

$$\overline{\phi_{1/2}(s)\phi_{1/2}(t)} = \min\{s,t\} - st$$
$$+ \min\left\{ s + \frac{\text{sign}(1/2 - s)}{2}, t + \frac{\text{sign}(1/2 - t)}{2} \right\}$$
$$- \left(s + \frac{\text{sign}(1/2 - s)}{2} \right)\left(t + \frac{\text{sign}(1/2 - t)}{2} \right). \quad (65)$$

The average of the last term in Eq. (62) requires the evaluation of $\overline{\phi_{1/2}^{(N)}(s)}$ that must be performed, as anticipated, including the corrections to the limiting Brownian bridge distribution. Introducing for the sake of brevity $\varsigma = s + \sigma(s, 1/2)$, we get

$$\Pr[\phi_{1/2}^{(N)}(s) \in d\phi] = \frac{\exp\left(-\frac{\phi^2}{2s(1-s)+2\varsigma(1-\varsigma)}\right) d\phi}{\sqrt{2\pi}\sqrt{s(1-s)+\varsigma(1-\varsigma)}}$$
$$\times \left[1 + \frac{s-\varsigma}{\sqrt{N}} \left(\frac{(s+\varsigma-1)^2}{(s(1-s)+\varsigma(1-\varsigma))^2}\phi \right. \right.$$
$$\left. \left. - \frac{1-(s-\varsigma)^2-3s(1-s)-3\varsigma(1-\varsigma)}{3(s(1-s)+\varsigma(1-\varsigma))^3}\phi^3 \right) + o\left(\frac{1}{\sqrt{N}}\right) \right], \quad (66)$$

that provides

$$\overline{\phi_{1/2}^{(N)}(s)} = -\frac{\sigma(s, 1/2)}{\sqrt{N}}. \quad (67)$$

Collecting the results above, we finally obtain

$$\varepsilon_N^{(p)} = \frac{1}{2^p}\left[1 + \frac{1}{3N}\frac{p(p-2)(p-4)}{p-3} \right] + o\left(\frac{1}{N}\right). \quad (68)$$

We verified the previous formula numerically. The numerical results show a good agreement with the theoretical prediction in Eq. (68), see Fig. 2b.

Given the optimal permutation π, such that $\pi(i) = i + k \pmod N$, the correlation function for the matching field

$$\mu_i = y_{\pi(i)} - x_i \xrightarrow[i=Ns+1/2]{N\to+\infty} \mu(s) = \begin{cases} \frac{1}{2} & \text{if } 0 \le s < \frac{1}{2} \\ -\frac{1}{2} & \text{if } \frac{1}{2} < s \le 1, \end{cases} \quad (69)$$

is easily calculated as

$$c_p(r) := \frac{1}{\mathcal{N}(r)} \iint_0^1 \overline{\mu(s)\mu(t)}\delta(|s-t|-r)\, d s\, d t$$
$$= \begin{cases} \frac{3}{4} - \frac{1}{2(1-r)} & \text{if } 0 \le r < \frac{1}{2} \\ -\frac{1}{4} & \text{if } \frac{1}{2} < r \le 1. \end{cases} \quad (70)$$

More interestingly, we introduce the correlation function for the field

$$\hat\mu_i := \sqrt{N}\left[y_{\pi(i)} - x_i - \frac{\text{sign}(N-k-i)}{2} \right]. \quad (71)$$

For $i = Ns + 1/2$, in the $N \to \infty$ limit keeping s fixed, we have

$$\hat\mu_{Ns+1/2} \xrightarrow{N\to+\infty} \phi_{1/2}(s) + \tau =: \hat\mu(s). \quad (72)$$

For $0 \leq r \leq 1$,

$$\hat{c}_p(r) := \iint\limits_0^1 \frac{\overline{\hat{\mu}(s)\hat{\mu}(t)}}{\mathcal{N}(r)} \delta\left(|s-t| - r\right) \mathrm{d}\,s\,\mathrm{d}\,t$$

$$= \iint\limits_0^1 \frac{\overline{\left(\tau + \phi_{1/2}(s)\right)\left(\tau + \phi_{1/2}(t)\right)}}{\mathcal{N}(r)} \delta\left(|s-t| - r\right) \mathrm{d}\,s\,\mathrm{d}\,t$$

$$= \begin{cases} \frac{p^2 - 6p + 10}{6(p-3)^2} + \frac{10r^2 - 30r + 17}{6(p-3)(1-r)}r - \frac{pr(1-r)}{p-3} & \text{if } 0 \leq r \leq 1/2, \\ \frac{p^2 - 3p + 1}{6(p-3)^2} - \frac{3p-4}{3(p-3)}r + \frac{3p-5}{3(p-3)}r^2 & \text{if } 1/2 < r \leq 1. \end{cases} \tag{73}$$

Observe that

$$\lim_{p \to -\infty} \hat{c}_p(r) = \frac{1}{6} - r(1-r), \tag{74}$$

that coincides with the correlation function for the assignment problem with $p = 2$ on the circumference [11, 18]. This fact is not a coincidence. Indeed, for $p \to -\infty$, we have that (see below)

$$\lim_{p \to -\infty} \tau = -\int_0^1 \phi_{1/2}(s)\,\mathrm{d}\,s. \tag{75}$$

By comparison with Eq. (78) below, it will be clear that $\hat{\mu}(s)$ for $p \to -\infty$ coincides with the solution of the assignment problem for $p = 2$ on the circumference in which one set of points is translated by $1/2$. In Fig. 1 we compare the predictions above for c_p and $\hat{c}_p$ with our numerical results.

3. Periodic boundary conditions

Corollary II.8 states that, in the case of periodic boundary conditions, both for $p > 1$ and for $p < 0$, the optimal matching can be found searching for the optimal permutation in the set $\mathscr{C}_N$. The calculation above for the assignment problem on the interval, however, has to be slightly modified, due to the fact that the cost function is replaced by the one in Eq. (29), whereas Eq. (40) still holds. In particular, the average optimal cost has the form given in Eq. (58), with $\sigma(s,t) \equiv \sigma(t) = t \in [0,1)$, value of the global shift depending on the specific instance and on the value of p. The optimal cost can be written, in the large N limit, as

$$\int_0^1 \left[\left|t + \frac{\phi_t(s)}{\sqrt{N}}\right| \pmod{1/2}\right]^p \mathrm{d}\,s, \tag{76}$$

for a certain value of t obtained by minimization.

a. The $p > 1$ case The $p > 1$ case has been analyzed in Refs. [11, 17, 18]. In this case at the leading order we obtain $t = 0$ and therefore $t = o(1)$. An explicit expression of t is known for $p = 2$ and $p \to +\infty$ only. In
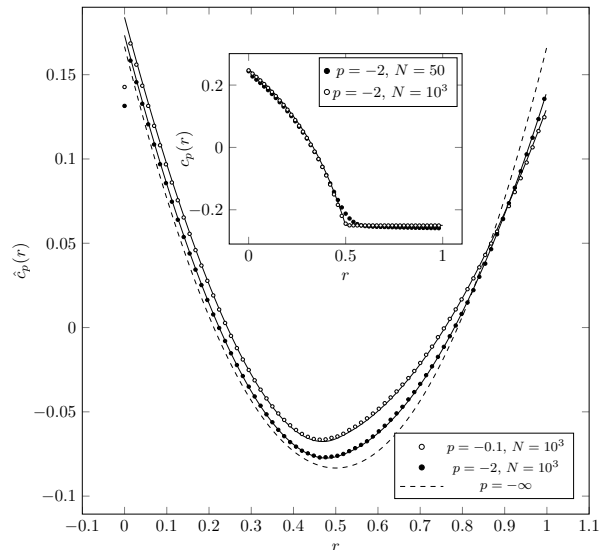

Figure 1. Correlation functions $c_p(r)$ (inset) and $\hat{c}_p(r)$ for the assignment problem on the interval Λ with $p < 0$, as defined in Eq. (70) and Eq. (73) respectively. Observe that a finite-size effect appears for $r \to 0$ in the numerical data for $\hat{c}_p(r)$. In all cases, the theoretical predictions are represented in solid line.

general, we can assume that $t = \tau/\sqrt{N} + o\left(1/\sqrt{N}\right)$. The value of τ is obtained minimizing, in the $N \to +\infty$ limit, the expression

$$\int_0^1 |\tau + \phi_0(s)|^p\,\mathrm{d}\,s. \tag{77}$$

For $p = 2$ we have [17, 18]

$$\varepsilon_N^{(2)} = \frac{1}{6N} + o\left(\frac{1}{N}\right), \tag{78a}$$

$$\tau = -\int_0^1 \phi_0(s)\,\mathrm{d}\,s. \tag{78b}$$

Unfortunately, no general expression for τ is available to our knowledge. Numerical simulations suggests that the average optimal cost scales as

$$\varepsilon_N^{(p)} = \frac{1}{N^{\frac{p}{2}}}\left[C_p + \frac{D_p}{N} + o\left(\frac{1}{N}\right)\right]. \tag{79}$$
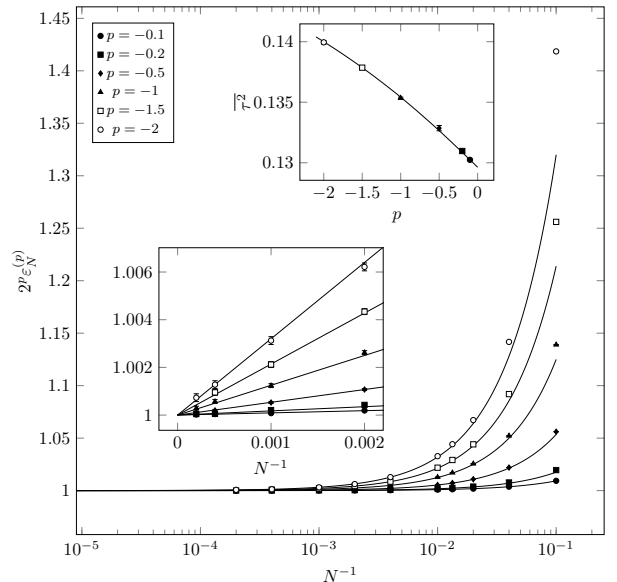
In Fig. 2c and in Table I we present our numerical results for the average optimal cost and its finite size corrections for the assigment problem on the circumference in the $p > 1$ case. The data have been obtained using Eq. (79) to extrapolate the $N \to +\infty$ limit for both C_p and D_p. In Ref. [12] the $p = 2$ case was carefully analyzed using a particular scaling ansatz, and the scaling in Eq. (79) was numerically verified. In particular, they obtained $C_2 = 0.166668(3)$ and $D_2 = -0.1645(13)$. We refer to Ref. [18] for further discussion on the correlation function on the circumference.
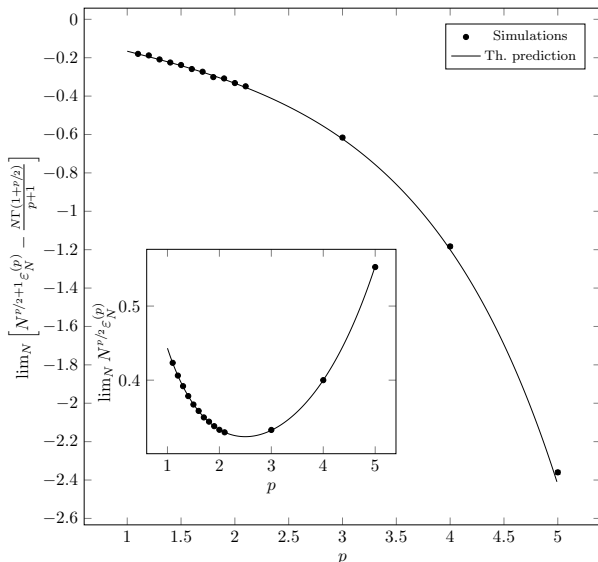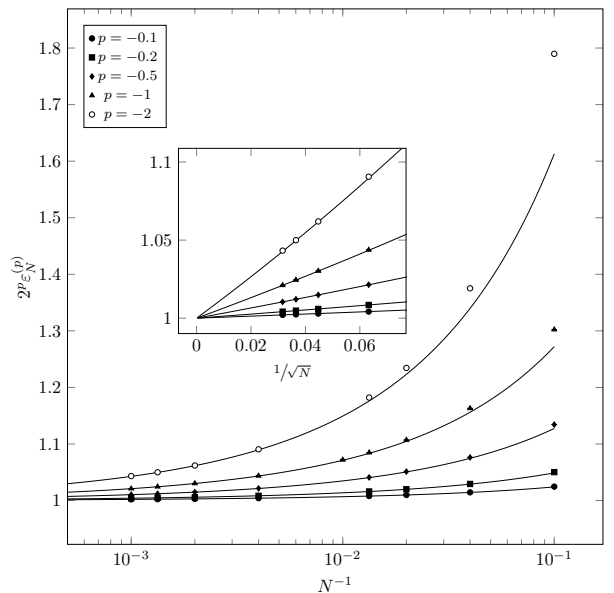
(a) Numerical results for the average optimal cost (inset) and its finite size corrections in the assignment problem with $p > 1$ on Λ. We fitted the obtained numerical results using the fitting function $f(N) = a + b/N$ for each value of p, a and b being fitting parameters corresponding to the leading cost and to the finite size corrections respectively. We present our numerical results and we compare them with our prediction given in Eq. (53).
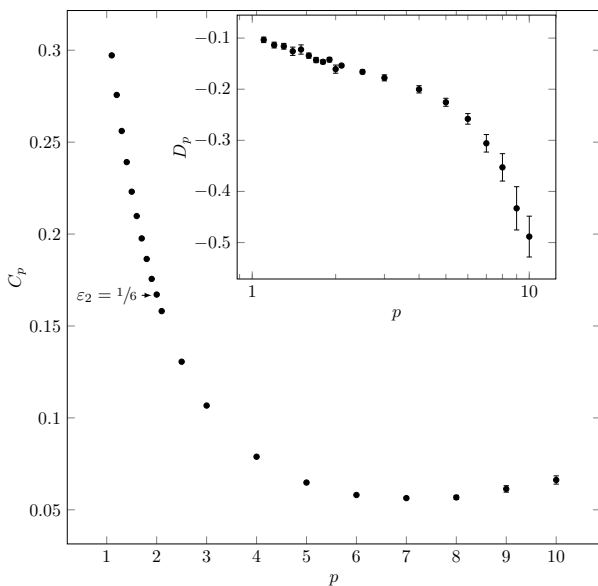
(b) Numerical results for the average optimal cost in the assignment problem on Λ with $p < 0$. The theoretical predictions are given in Eq. (68) and they are represented in solid lines. In the upper inset, numerical results for $\overline{\tau^2}$ for different values of p, obtained using two-parameters fitting function in the form $f(N) = \alpha + \beta/N$, α being the numerical estimation for $\overline{\tau^2}$. We compare them with the prediction in Eq. (64).

(c) Numerical results for the average optimal cost (inset) and its finite size corrections in the assignment problem with $p > 1$ on the circumference. We fitted the numerical results obtained for different values of N at fixed p using the fitting function in the form given in Eq. (79). See also Table I.

(d) Numerical results for the average optimal cost and its finite size corrections in the assignment problem with $p < 0$ on the circumference. We present our numerical results and we compare them with our theoretical predictions for the asymptotic behavior given in Eq. (82) (solid lines).

Figure 2. Numerical results for the assignment problem. Error bars are typically smaller than the markers in the figures.

b. *The $p < 0$ case* For $p < 0$ a more detailed computation can be performed. At the leading order, the minimum is obtained for $t = 1/2 + O\left(1/\sqrt{N}\right)$, as in the case of open boundary condition. Under the assumption

$t = 1/2 + \tau/\sqrt{N}$ we obtain

p	C_p	D_p
1.1	0.2972(2)	-0.103(5)
1.2	0.2756(2)	-0.114(6)
1.3	0.2561(3)	-0.116(6)
1.4	0.2392(3)	-0.126(8)
1.5	0.2231(4)	-0.122(9)
1.6	0.2098(2)	-0.134(5)
1.7	0.1976(2)	-0.143(5)
1.8	0.1864(2)	-0.146(4)
1.9	0.1756(2)	-0.142(4)
2.0	0.1671(3)	-0.161(8)
2.1	0.1580(1)	-0.154(3)
2.5	0.1305(2)	-0.166(4)
3.0	0.1067(2)	-0.178(6)
4.0	0.0788(3)	-0.200(7)
5.0	0.0648(3)	-0.226(7)
6.0	0.0580(4)	-0.26(1)
7.0	0.0563(7)	-0.31(2)
8.0	0.057(1)	-0.35(2)
9.0	0.061(2)	-0.43(4)
10	0.066(2)	-0.489(4)

Table I. Numerical results for the average optimal cost and its finite size corrections in the assignment problem on the circumference. For each value of p, the average optimal cost has been evaluated averaging over at least 10^4 instances with size between $N = 10$ and $N = 2.5 \cdot 10^3$. Subsequently, a fit has been performed using Eq. (79) to extract C_p and D_p.

$$
\int_0^1 \left[\frac{1}{2} - \frac{\left| \tau + \phi_{1/2+\tau/\sqrt{N}}^{(N)}(s) \right|}{\sqrt{N}} \right]^p \mathrm{d}s =
$$

$$
= \frac{1}{2^p} - \frac{p}{2^{p-1}\sqrt{N}} \int_0^1 \left| \tau + \phi_{1/2}(s) \right| \mathrm{d}s
$$

$$
- \frac{p}{2^{p-1}\sqrt{N}} \int_0^1 \left(\phi_{1/2}^{(N)}(s) - \phi_{1/2}(s) + \frac{\tau \, \partial_t \phi_t(s)|_{t=1/2}}{\sqrt{N}} \right) \mathrm{d}s
$$

$$
+ \frac{p(p-1)}{2^{p-1}N} \int_0^1 \left(\tau + \phi_{1/2}(s) \right)^2 \mathrm{d}s + o\left(\frac{1}{N} \right). \quad (80)
$$

In the expression above we took into account that $\phi_{1/2}^{(N)}(s) - \phi_{1/2}(s)$ is infinitesimal quantity for large N, due to Eq. (35). Moreover, $\int_0^1 \partial_t \phi_t(s) \, \mathrm{d}s = \partial_t \int_0^1 \phi_t(s) \, \mathrm{d}s = 0$, see Eq. (57). We have then that the third contribution in the previous equation is $O(1/N)$. Minimizing respect to τ, we obtain. up to higher order terms,

$$
\int_0^1 \mathrm{sign}\left(\tau + \phi_{1/2}(s) \right) \mathrm{d}s = 0, \quad (81)
$$

and therefore the optimal value τ_ϕ of τ depends on the instance $\phi_{1/2}$, i.e., on the properties of the Brownian bridge process, and not on p. The average optimal cost is

$$
\varepsilon_N^{(p)} = \frac{1}{2^p} - \frac{p\lambda_1}{2^{p-1}\sqrt{N}} + \frac{p(p-1)\lambda_2}{2^{p-1}N} + o\left(\frac{1}{N} \right). \quad (82a)
$$

where the quantities

$$
\lambda_1 := \overline{\int_0^1 \left| \tau_\phi + \phi_{1/2}(s) \right| \mathrm{d}s} = 0.3217(5), \quad (82b)
$$

$$
\lambda_2 := \overline{\int_0^1 \left(\tau_\phi + \phi_{1/2}(s) \right)^2 \mathrm{d}s} = 0.1717(5), \quad (82c)
$$

are fixed numbers related to the Brownian bridge process only, that we evaluated numerically. We numerically verified Eq. (82). Our numerical results are given in Fig. 2d and they show a good agreement with the theoretical prediction.

III. THE RANDOM EUCLIDEAN MATCHING PROBLEM

In the *random Euclidean matching problem in one dimension* we associate to the set of $2N$ vertices of the complete graph $\mathcal{K}_{2N}$ a set of $2N$ points $\Xi_N := \{x_i\}_{i=1,\dots,2N}$ independently and randomly generated on Λ with uniform distribution. Again, we will assume that the points are labeled in such a way that $0 \leq x_1 < x_2 < \cdots < x_{2N} \leq 1$. In this case, a matching μ is any partition of Ξ_{2N} in subsets of two elements only, its cardinality being N. We will consider the following *matching cost* associated to μ,

$$
\varepsilon_N^{(p)}(\mu) := \frac{1}{N} \sum_{(i,j) \in \mu} |x_i - x_j|^p, \quad p \in \mathbb{R}. \quad (83)
$$

As in the assignment problem, we are interested in the average optimal cost

$$
\varepsilon_N^{(p)} := \overline{\min_\mu \varepsilon_N^{(p)}(\mu)}, \quad (84)
$$

and in its asymptotic behavior for $N \to +\infty$.

A. Open boundary conditions

For $p > 1$, the optimal solution on the interval has a simple structure. In particular, the couple (x_i, x_j), $i < j$, belongs to the optimal matching if, and only if, i is odd and $j = i + 1$. This statement follows directly from the direct inspection of the $N = 2$ case. We have indeed that given the generic configuration
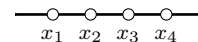
the minimum cost configuration has always the structure
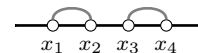
The study of the properties of the optimal matching is reduced therefore to the study of *spacings* between successive random points on Λ. The optimal cost for $p > 1$ is therefore given by

$$\min_{\mu} \varepsilon_N^{(p)}(\mu) = \frac{1}{N} \sum_{i=1}^{N} \varphi_{2i-1}^p, \quad \varphi_i := x_{i+1} - x_i. \quad (85)$$

Let us first observe that the distribution of the *ordered* set $\mathbf{x} = (x_1, \ldots, x_{2N})$ is given by

$$\rho_N(\mathbf{x}) = (2N)! \prod_{i=0}^{2N} \theta(x_{i+1} - x_i), \quad x_0 \equiv 0, \ x_{2N+1} \equiv 1. \quad (86)$$

It follows that

$$\varrho_N(\varphi_0, \ldots, \varphi_{2N}) = (2N)! \prod_{i=0}^{2N} \theta(\varphi_i) \quad \text{for} \ \sum_{i=0}^{2N} \varphi_i = 1. \quad (87)$$

In particular, this implies that for the spacing φ_l we have

$$\varrho_N^{(1)}(\varphi_l) = (2N)! \left[\prod_{\substack{k=0 \\ k \neq l}}^{2N} \int_0^{+\infty} d\varphi_k \right] \delta\left(\sum_{j=0}^{2N} \varphi_j - 1 \right)$$

$$= (2N)! i^{2N} \lim_{\epsilon \to 0^+} e^{\epsilon(1-\varphi_l)} \int_{-\infty}^{+\infty} \frac{e^{-i\lambda(1-\varphi_l)}}{(\lambda + i\epsilon)^{2N}} \frac{d\lambda}{2\pi}$$

$$= \begin{cases} 2N(1-\varphi_l)^{2N-1} & \text{for } 0 < \varphi_l < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (88)$$

Observe that the shape of the distribution is not dependent on l. Moreover,

$$\varepsilon_N^{(p)} \equiv \overline{\varphi_l^p} = \frac{\Gamma(2N+1)\Gamma(1+p)}{\Gamma(2N+1+p)}$$

$$= \frac{1}{N^p} \frac{\Gamma(p+1)}{2^p} \left[1 - \frac{p(p+1)}{4N} + o\left(\frac{1}{N}\right) \right]. \quad (89)$$

The joint density distribution $\varrho^{(2)}$ of the couple (φ_i, φ_j), $i \neq j$, can be similarly evaluated. As proven, for example, in Ref. [27], we have that

$$\varrho_N^{(2)}(\varphi_i, \varphi_j) =$$
$$= 2N(2N-1)(1-\varphi_i-\varphi_j)^{2(N-1)}\theta(\varphi_i)\theta(\varphi_j)\theta(1-\varphi_i-\varphi_j), \quad (90)$$

implying

$$\overline{\varphi_i \varphi_j} = \frac{1}{(2N+1)(2N+2)}. \quad (91)$$

Observe once again that no dependence on i and j appears on the right hand side of the previous equations. It is clear that in this case $\varphi_i \sim N^{-1}$. We introduce the

rescaled variables $\phi_i = 2N\varphi_i$, whose asymptotic distribution, for $0 \leq \phi_i \leq 2N$, is given by

$$\hat{\varrho}_N^{(1)}(\phi_i) = e^{-\phi_i} \left[1 - \frac{\phi_i(\phi_i - 2)}{4N} + o\left(\frac{1}{N}\right) \right]. \quad (92a)$$

Similarly, the joint probability distribution for $\phi_i, \phi_j \geq 0$, $0 \leq \phi_i + \phi_j \leq 2N$, is

$$\hat{\varrho}_N^{(2)}(\phi_i, \phi_j) =$$
$$= e^{-\phi_i - \phi_j} \left[1 - \frac{(\phi_i+\phi_j)^2 - 4(\phi_i+\phi_j)+2}{4N} + o\left(\frac{1}{N}\right) \right]. \quad (92b)$$
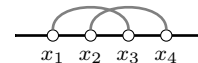
We can therefore write

$$\overline{\phi_i} = 1 - \frac{1}{2N} + o\left(\frac{1}{N}\right), \quad (93a)$$

$$\overline{\phi_i \phi_j} = 1 - \frac{3}{2N} + o\left(\frac{1}{N}\right). \quad (93b)$$

This implies

$$\lim_N \left(\overline{\phi_i \phi_j} - \overline{\phi_i}\,\overline{\phi_j} \right) = 0. \quad (94)$$

For $p < 0$ it is easily seen that, for $N = 2$, the optimal solution is always the crossing one, i.e., in the form

This can be proved again by direct inspection, in the spirit of the analysis in Proposition II.2 and observing that a crossing solution is always possible. It follows that the optimal matching on a set of $2N$ points on the interval is given by the set of couples $\{(x_i, x_{i+N})\}_{i=1,\ldots,N}$, such that, in the pictorial representation above, each arc corresponding to a matched coupled crosses all the remaining $N - 1$ arcs. The optimal cost per edge is

$$\min_{\mu} \varepsilon_N^{(p)}(\mu) = \frac{1}{N} \sum_{i=1}^{N} (x_{i+N} - x_i)^p. \quad (95)$$

The analysis proceed as in the $p > 1$ case. To evaluate the average optimal cost, denoting by $\varphi_l := x_{l+N} - x_l$, we have

$$\Pr[\varphi_l \in d\varphi] = B(N; 2N, \varphi)\theta(\varphi)\theta(1-\varphi)\frac{N\,d\varphi}{\varphi} \quad (96)$$

that is the probability that given $2N$ points at random N of them are in an interval of length φ. Of course the distribution of φ_l does not depend on l. It follows that, for any real γ such that $N > -\gamma$,

$$\overline{\varphi_l^\gamma} = \frac{\Gamma(2N+1)\Gamma(N+\gamma)}{\Gamma(N)\Gamma(2N+\gamma+1)} = \frac{1}{2^\gamma} + \frac{\gamma}{2^{\gamma+2}}\frac{\gamma-3}{N} + o\left(\frac{1}{N}\right), \quad (97)$$

and therefore, for $p < 0$ and $N > -p$,

$$\varepsilon_N^{(p)} = \frac{\Gamma(2N+1)\Gamma(N+p)}{\Gamma(N)\Gamma(2N+p+1)} = \frac{1}{2^p} + \frac{p(p-3)}{2^{p+2}N} + o\left(\frac{1}{N}\right). \tag{98}$$

For $N \gg 1$ expectation values in the distribution given by Eq. (96) can be evaluated by the saddle-point method. By performing the shift around the saddle point value

$$\varphi_l = \frac{1}{2} + \frac{1}{2\sqrt{N}}\phi_l \tag{99}$$

we recover the distribution for ϕ_l as

$$\Pr[\phi_l \in \mathrm{d}\,\phi] \simeq$$
$$\simeq \frac{\mathrm{e}^{-\phi^2}}{\sqrt{\pi}}\left[1 - \frac{1}{\sqrt{N}}\phi - \frac{4\phi^4 - 8\phi^2 + 1}{8N} + o\left(\frac{1}{N}\right)\right]\mathrm{d}\,\phi. \tag{100}$$

For example the evaluation of

$$\overline{\varphi_l^{\gamma}} = \frac{1}{2^{\gamma}}\overline{\left(1 + \frac{1}{\sqrt{N}}\phi_l\right)^{\gamma}}$$
$$\simeq \frac{1}{2^{\gamma}}\left[1 + \frac{\gamma}{\sqrt{N}}\overline{\phi_l} + \frac{\gamma(\gamma-1)}{2N}\overline{\phi_l^2} + o\left(\frac{1}{N}\right)\right] \tag{101}$$

will provide the result given in Eq. (97), because

$$\overline{\phi_l} = -\frac{1}{2\sqrt{N}} + o\left(\frac{1}{N}\right) \tag{102a}$$

$$\overline{\phi_l^2} = \frac{1}{2} - \frac{1}{4N} + o\left(\frac{1}{N}\right). \tag{102b}$$

For the evaluation of correlations, we have, for $0 \le l < k \le N$,

$$\Pr[x_l \in \mathrm{d}\,x, x_{l+N} \in \mathrm{d}\,x', x_k \in \mathrm{d}\,y, x_{k+N} \in \mathrm{d}\,y'] =$$
$$= \mathrm{d}\,x\,\mathrm{d}\,x'\,\mathrm{d}\,y\,\mathrm{d}\,y'(2N)!\frac{x^{l-1}(y-x)^{k-l-1}(x'-y)^{N+l-k-1}}{\Gamma(l)\Gamma(k-l)\Gamma(N+l-k)}$$
$$\times \frac{(y'-x')^{k-l-1}(1-y')^{N-k}}{\Gamma(k-l)\Gamma(N-k+1)}$$
$$\times \theta(x)\theta(y-x)\theta(x'-y)\theta(y'-x')\theta(1-y'). \tag{103}$$

We get therefore

$$\overline{\varphi_l\varphi_k} = \frac{N(N+1) - |k-l|}{(2N+2)(2N+1)}. \tag{104}$$

Eq. (104) suggests the introduction of the variables x, y, such that $Nx = l$, $Ny = k$, and, with reference to Eq. (99), of the field variable $\phi(x) := \phi_{Nx}$. We have that in the large N limit

$$\overline{\phi(x)\phi(y)} - \overline{\phi(x)}\,\overline{\phi(y)} =$$
$$= \frac{1 - 2|y-x|}{2} + \frac{3|y-x|}{2N} + O\left(\frac{1}{N^2}\right). \tag{105}$$

In Fig. 3a and Fig. 3b we compare our theoretical results with the output of numerical simulations for the $p > 1$ and the $p < 0$ case respectively.

B. Periodic boundary condition

The case of periodic boundary condition for $p > 1$ can be easily obtained. Observe indeed that, in the case of $2N = 4$ points on the circumference, given the crossing matching, we can always lower the cost considering one of the noncrossing solutions, i.e.,
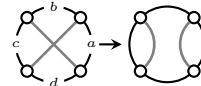
where the arrows denote the transition from a given matching to another one with lower cost. Indeed, with reference to the figure above, denoting by a, b, c and d the length of the four arcs with extremes the considered points, let us suppose, without loss of generality, that $a+b \le 1/2$ and $b+c \le 1/2$. Then, $a^p+c^p < (a+b)^p+(b+c)^p$ for $p > 1$. Moreover, given the matching
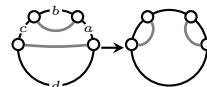
if $a + b + c \le 1/2$, then $a^p + c^p \le (a + b + c)^p + b^p$, i.e., there are no nested matchings in a half-circumference. Applying these rules iteratively to the case of $2N$ points on the circumference, we find that, ordering the points according to a reference orientation on the circumference, we have two possible optimal matching configurations, namely, for $i = 1, \ldots, N$, the $2i$-th point is associated either to the $(2i + 1 \pmod{2N})$-th point, or to the $(2i - 1 \pmod{2N})$-th point. Pictorially,

$$\tag{106}$$

The distribution of $2N$ spacings $\{\varphi_i\}$ generated by $2N$ random points on the circumference is given by

$$\varrho(\varphi_1, \ldots \varphi_{2N}) = \Gamma(2N)\delta\left(\sum_{i=1}^{2N}\varphi_i - 1\right)\prod_{i=1}^{N}\theta(\varphi_i). \tag{107}$$

We assume here that we choose one of the points as origin, and an orientation on the circumference, such that the intervals φ_i are labelled accordingly. Let $p(\Phi_1, \ldots, \Phi_{2N})$ be the probability for the quantities $\Phi_i := N^p\varphi_i^p$, that can be straightforwardly obtained from Eq. (107). The variables $\{\Phi_i\}_i$ have mean

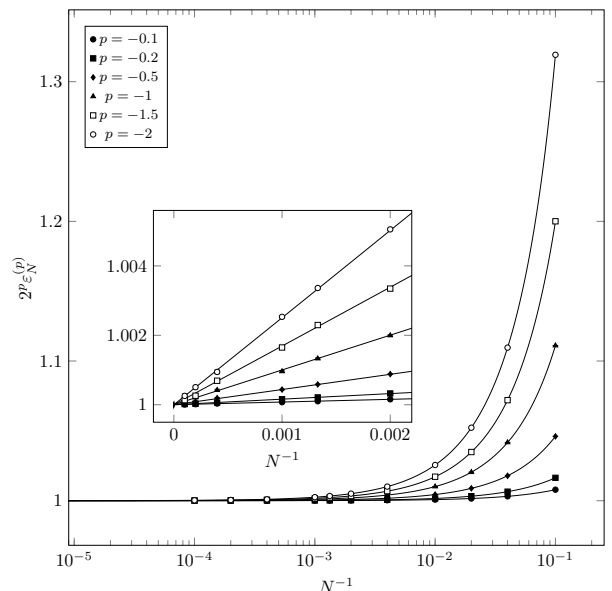$$\mu := \overline{\Phi_i} = N^p\frac{\Gamma(2N)\Gamma(p+1)}{\Gamma(2N+p)}, \tag{108a}$$

and variance

$$\overline{(\Phi_i - \mu)^2} =$$
$$= N^{2p}\left[\frac{\Gamma(2N)\Gamma(1+2p)}{\Gamma(2N+2p)} - \frac{\Gamma^2(2N)\Gamma^2(1+p)}{\Gamma^2(2N+p)}\right]$$
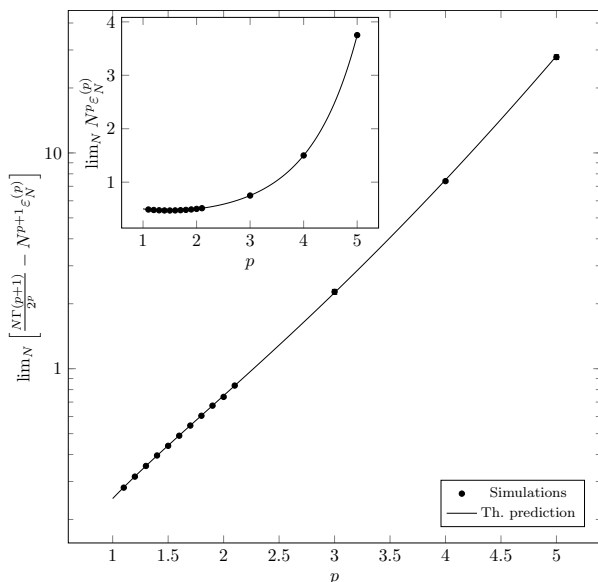
(a) Numerical results for the finite size corrections to the average optimal cost (inset) in the matching problem with $p > 1$ on Λ. We compare our numerical results with our numerical prediction given in Eq. (89) (solid lines).

(b) Numerical results for average optimal cost in the matching problem on Λ with $p < 0$. The theoretical predictions (solid lines), for any value of $N > -p$, are given by Eq. (98). Observe that for $N \gg 1$ $2^p \varepsilon_N^{(p)}$ is linear in N^{-1} (inset).
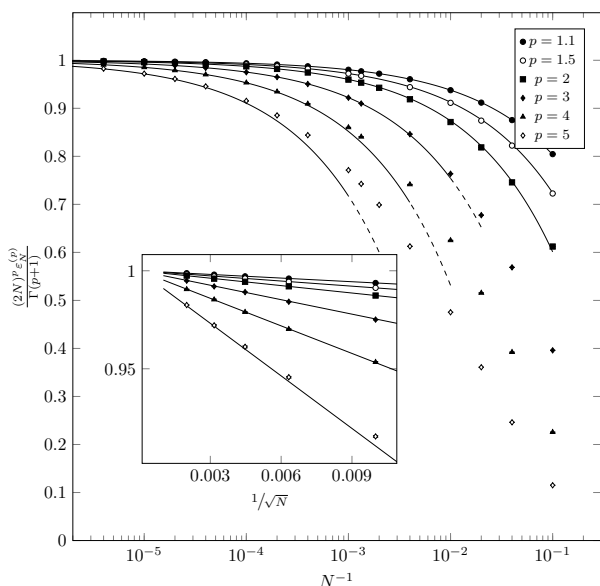
(c) Numerical results for average optimal cost in the matching problem on the circumference with $p > 1$. The theoretical predictions (solid lines) for the asymptotic behavior are given by Eq. (110). Observe that for $N \gg 1$ the corrections to the asymptotic cost scale as $1/\sqrt{N}$ (inset).
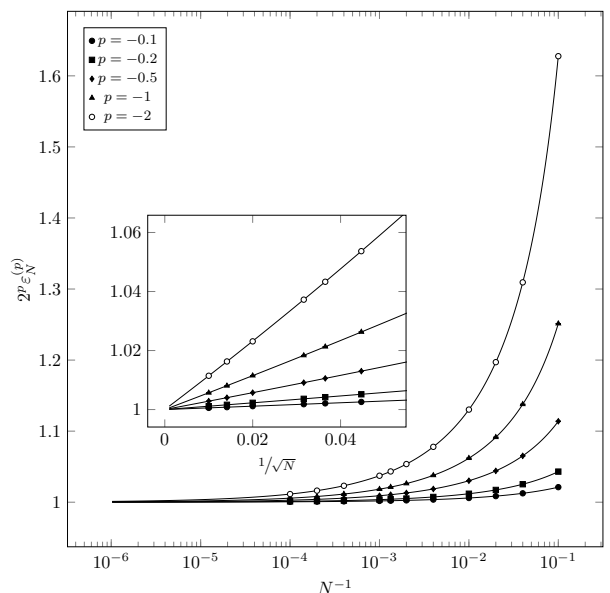
(d) Numerical results for average optimal cost in the matching problem on the circumference with $p < 0$. The theoretical predictions (solid lines) are given by Eq. (112) and they are correct for all values of $N > -p$, as expected. Observe that for $N \gg 1$ $2^p \varepsilon_N^{(p)}$ is linear in $1/\sqrt{N}$ (inset).

Figure 3. Numerical results for the matching problem. Error bars are typically smaller than the markers in the figures.

$$= \frac{\Gamma(2p+1) - \Gamma^2(p+1)}{2^{2p}} + o(1) \equiv \sigma^2 + o(1). \quad (108b)$$

The variables are however not independent, due to the overall constraint $\sum_i \Phi_i^{1/p} = N$. We have that, for $i \neq j$,

$$\overline{(\Phi_i - \mu)(\Phi_j - \mu)} =$$
$$= N^{2p}\Gamma^2(1+p)\Gamma(2N)\left[\frac{1}{\Gamma(2N+2p)} - \frac{\Gamma(2N)}{\Gamma^2(2N+p)}\right]$$
$$= -\frac{p^2\Gamma^2(1+p)}{2^{2p+1}N} + o\left(\frac{1}{N}\right) = \frac{\rho}{N} + o\left(\frac{1}{N}\right). \quad (108c)$$

The optimal cost in the matching problem on the circumference with $p > 1$ is given by

$$\varepsilon_N^{(p)} = \overline{\min\left\{ \frac{1}{N^{1+p}} \sum_{i=1}^N \Phi_{2i-1}, \frac{1}{N^{1+p}} \sum_{i=1}^N \Phi_{2i} \right\}}. \quad (109)$$

Using the results given in Appendix C, we have in this case that

$$(2N)^p \varepsilon_N^{(p)} =$$
$$= \Gamma(p+1) + \sqrt{\frac{\Gamma(2p+1) - \Gamma^2(p+1)}{\pi N}} + o\left(\frac{1}{\sqrt{N}}\right). \quad (110)$$

In Fig. 3c we compare our numerical results with the theoretical prediction in Eq. (110).

In the $p < 0$ case, as in the case of open boundary conditions, we have that, given four points on the circumference, the optimal solution is always the crossing one. Let us consider indeed
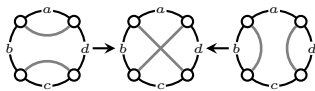
and let us assume, without loss of generality, that $a + b \leq 1/2$ and $b + c \leq 1/2$. We have that $a^p + c^p \geq (a+b)^p + (b+c)^p$. With reference to the figure above, if $d \geq 1/2$ we also have $b^p + (a+b+c)^p > (a+b)^p + (b+c)^p$, where we have used the fact that $f(x,a) = x^p - (x+a)^p$ is a decreasing function for $x > 0$ and $a > 0$. If $d = 1 - a - b - c \leq 1/2$, then $b \leq d$ and therefore we have $b^p + d^p \geq 2d^p \geq (a+b)^p + (b+c)^p$. This fact implies that the minimum cost matching is obtained coupling the ith point to the $i + N \pmod{2N}$ point on the circumference, where the points are supposed to be ordered according to a reference orientation on the circumference. For example, we will have that
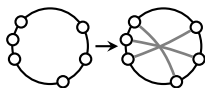
To find the average optimal cost, observe that, fixing the origin of our reference system in the point i, the distance of the point $i + N \pmod{N}$ from the ith point on the circumference is distributed as

$$\Pr[\varphi_i \in \mathrm{d}\,\varphi] = N\theta(\varphi)\theta\left(\frac{1}{2} - \varphi\right)\mathrm{d}\,\varphi \times$$
$$\times \left[\frac{B(N; 2N, \varphi)}{\varphi} + \frac{B(N; 2N, 1 - \varphi)}{1 - \varphi}\right]. \quad (111)$$

As in the case of periodic boundary conditions, the previous distribution does not depend on i. We obtain, for $N > -p$, the average optimal cost straightforwardly as

$$\varepsilon_N^{(p)} = N\binom{2N}{N}\beta_{1/2}(N + p, N)$$

$$= \frac{1}{2^p}\left[1 - \frac{p}{\sqrt{\pi N}} + \frac{p(p-1)}{4N} + o\left(\frac{1}{N}\right)\right], \quad (112)$$

where we have introduced the incomplete Beta function

$$\beta_s(a, b) := \int_0^s t^{a-1}(1-t)^{b-1}\,\mathrm{d}\,t. \quad (113)$$

In Fig. 3d we show that the results of our numerical simulations are in agreement with Eq. (112).

IV. CONCLUSIONS

In the present paper we discussed the Euclidean matching problem and the Euclidean assignment problem on a set of $2N$ points both on the line and on the circumference.

We first stated some fundamental properties of the Euclidean assignment problem on the line for a large class of cost functions $c(z)$, that we called $\mathscr{C}$-functions, and for strictly increasing cost functions. We proved that, for these classes of cost functions, the optimal matching $x_i \to y_{\pi(i)}$ between the set of points $0 \leq x_1 < x_2 < \cdots < x_N \leq 1$ and the set of points $0 \leq y_1 < y_2 < \cdots < y_N \leq 1$ can be expressed as a permutation in the form $\pi(i) = i + k \pmod{N}$ for some k, in the case of strictly increasing cost functions the optimal permutation being the identical permutation, $\pi(i) = i$. We considered then the assignment problem both on the line and on the circumference in presence of disorder, assuming the points to be uniformly and randomly generated on the considered domain. We chose the cost function $c(z) = z^p$ with $p \in \mathbb{R} \setminus [0, 1]$, that is a $\mathscr{C}$-function for $p < 0$ and a strictly increasing function for $p > 1$. The analytical investigation allowed us to relate the optimal solution, in all the considered cases, to a well known Gaussian stochastic process, namely the Brownian bridge process, in the $N \to +\infty$ limit. Then, we analytically derived the expression for the average optimal cost and its finite size corrections for the considered range of values of p, and we gave an explicit expression of the correlation functions for the optimal solutions. In particular, we computed the constants appearing in the average cost on the unit interval

$$\varepsilon_N^{(p)} = \begin{cases} \frac{\Gamma(1+p/2)}{p+1}\left[1 - \frac{1}{N}\frac{p(p+2)}{8} + o\left(\frac{1}{N}\right)\right]\frac{1}{N^{p/2}} & \text{for } p > 1, \\ \frac{1}{2^p}\left[1 + \frac{1}{N}\frac{p(p-2)(p-4)}{3(p-3)} + o\left(\frac{1}{N}\right)\right] & \text{for } p < 0, \end{cases} \quad (114)$$

and those on the unit circumference where

$$\varepsilon_N^{(p)} = \begin{cases} \left[C_p + \frac{D_p}{N} + o\left(\frac{1}{N}\right)\right]\frac{1}{N^{p/2}} & \text{for } p > 1, \\ \frac{1}{2^p}\left[1 - \frac{2p\lambda_1}{\sqrt{N}} + \frac{2p(p-1)\lambda_2}{N} + o\left(\frac{1}{N}\right)\right] & \text{for } p < 0, \end{cases} \quad (115)$$

where the constants λ_1 and λ_2 were defined in Eq. (82b) and Eq. (82c) respectively. Unfortunately, in the $p > 1$ case, only C_2 is known analytically.

We analyzed in a similar way the Euclidean matching problem. In particular, for the average cost on the unit interval we computed the constants appearing in the expansions

$$\varepsilon_N^{(p)} = \begin{cases} \frac{\Gamma(p+1)}{2^p}\left[1 - \frac{p(p+1)}{4N} + o\left(\frac{1}{N^2}\right)\right]\frac{1}{N^p} & \text{for } p > 1, \\ \frac{1}{2^p}\left[1 + \frac{p(p-3)}{4N} + o\left(\frac{1}{N^2}\right)\right] & \text{for } p < 0, \end{cases} \tag{116}$$

and those for the problem on the unit circumference where

$$\varepsilon_N^{(p)} = \begin{cases} \frac{\Gamma(p+1)}{2^p}\left[1 + \sqrt{\left(\frac{\Gamma(2p+1)}{\Gamma^2(p+1)} - 1\right)\frac{1}{\pi N}} + o\left(\frac{1}{\sqrt{N}}\right)\right]\frac{1}{N^p} \\ \hspace{5cm} \text{for } p > 1, \\ \frac{1}{2^p}\left[1 - \frac{p}{\sqrt{\pi N}} + \frac{p(p-1)}{2^{p+2}N} + o\left(\frac{1}{N}\right)\right] \\ \hspace{5cm} \text{for } p < 0. \end{cases} \tag{117}$$

The first remark is the different leading power of N appearing here, that is N^{-p}, at variance with the assignment case where it was $N^{-p/2}$. Second, we observe that, both in the case of the assignment problem with $p < 0$ and in the case of the matching problem, the finite-size corrections to the average optimal cost change their scaling properties when open boundary conditions are replaced by periodic boundary conditions, i.e., when we consider the problem on the circumference instead of the interval. In particular, in the case of open boundary conditions the finite-size corrections scale as $O\left(1/N\right)$, whereas in the case of periodic boundary conditions, they scale as $O\left(1/\sqrt{N}\right)$. This fact can be observed both in Fig. 2 and in Fig. 3.

V. ACKNOWLEDGMENTS

Appendix A: On the convexity property of $\mathscr{C}$-functions

In this appendix we show that Eq. (15a) is equivalent to convexity if the function f is continuous on the interval $(0, 1)$. Introducing

$$R(z_1, z_2) \coloneqq \frac{f(z_1) - f(z_2)}{z_1 - z_2}, \tag{A1}$$

Eq. (15a) can be written as

$$R(z_1, z_2) \leq R(z_1 + \eta, z_2 + \eta). \tag{A2}$$

Observe that $R(z_1, z_2)$ is a symmetric function of its arguments, and, moreover, for $n \in \mathbb{N}$, we can write

$$R(z_1, z_2) =$$
$$= \frac{1}{n}\sum_{k=0}^{n-1} R\left(z_1 + k\frac{z_2 - z_1}{n}, z_2 + (k+1)\frac{z_2 - z_1}{n}\right)$$
$$\leq R\left(z_1, z_2 + \frac{z_2 - z_1}{n}\right), \tag{A3a}$$

and

$$R\left(z_1, z_2 + n(z_2 - z_1)\right) =$$
$$= \frac{1}{n}\sum_{k=0}^{n-1} R\left(z_1 + k(z_2 - z_1), z_2 + (k+1)(z_2 - z_1)\right)$$
$$\geq R\left(z_1, z_2\right). \tag{A3b}$$

Eqs. (A3) are equivalent to say that the function $R(z_1, z_2)$ is monotonically non decreasing respect to each one of its arguments taking the other one fixed, provided that the ratio of the considered intervals is rational. If the function is continuous, we can extend this property to an arbitrary couple of intervals, and therefore, for $\eta > 0$, we can simply state the stronger chain of inequalities

$$R(z_1, z_2) \leq R(z_1, z_2 + \eta) \leq R(z_1 + \eta, z_2 + \eta), \tag{A4}$$

that is R is monotonically non decreasing respect to each of its arguments taking the other one fixed. This property is equivalent to convexity. Indeed, if we consider $\xi = tx_1 + (1-t)x_2$ with $t \in [0, 1]$ and $x_1 < x_2$,

$$R(\xi, x_1) \leq R(\xi, x_2), \tag{A5a}$$
$$\frac{f(\xi) - f(x_1)}{(1-t)(x_2 - x_1)} \leq \frac{f(x_2) - f(\xi)}{t(x_2 - x_1)}, \tag{A5b}$$
$$f(\xi) \leq tf(x_1) + (1-t)f(x_2). \tag{A5c}$$

Appendix B: Derivation of Eq. (52)

In this appendix we will sketch the derivation of Eq. (52). We have to evaluate

$$\Pr[\varphi_k \in \mathrm{d}\,\varphi] = \mathrm{d}\,\varphi\binom{N}{k}^2 k^2 \iint_0^1 \delta\left(\varphi - y + x\right) \times$$
$$(xy)^{k-1}[(1-x)(1-y)]^{N-k}\,\mathrm{d}\,x\,\mathrm{d}\,y \tag{B1}$$

for $N \gg 1$. Let us write now, for $N \gg 1$,

$$k = Ns + \frac{1}{2}, \quad s \in (0, 1). \tag{B2}$$

The integral above can be written as

$$\frac{\Gamma^2(N+1)}{\Gamma^2(Ns+1/2)\Gamma^2(N+1/2-Ns)} \iint_0^1 \frac{\delta\left(\varphi - y + x\right)}{\sqrt{xy(1-x)(1-y)}}$$
$$\times \mathrm{e}^{Ns\ln(xy) + N(1-s)\ln[(1-x)(1-y)]}\,\mathrm{d}\,x\,\mathrm{d}\,y. \tag{B3}$$

We will evaluate the integral above using the saddle point method. In particular, the saddle point $(x_{\rm sp}, y_{\rm sp})$ is obtained for

$$x_{\rm sp} = y_{\rm sp} = s. \tag{B4}$$

Observe now that, at fixed s, for $N \gg 1$,

$$\frac{\Gamma(N+1)}{\Gamma(Ns+{}^1\!/_2)\Gamma(N+{}^1\!/_2-Ns)} =$$

$$= \sqrt{N}\frac{\mathrm{e}^{-N[s\ln s + (1-s)\ln(1-s)]}}{\sqrt{2\pi}}$$

$$\times \left[1 + \frac{1+2s(1-s)}{24Ns(1-s)} + o\left(\frac{1}{N}\right)\right], \tag{B5}$$

where we have used the Stirling expansion for $n \gg 1$

$$n! = \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n\left[1 + \frac{1}{12n} + o\left(\frac{1}{N}\right)\right]. \tag{B6}$$

We have therefore

$$\frac{\Gamma^2(N+1)}{\Gamma^2(Ns+{}^1\!/_2)\Gamma^2(N+{}^1\!/_2-Ns)}\int\limits_0^1\!\!\int \frac{\delta(\varphi-y+x)}{\sqrt{xy(1-x)(1-y)}}\,\mathrm{e}^{Ns\ln(xy)+N(1-s)\ln[(1-x)(1-y)]}\,\mathrm{d}\,x\,\mathrm{d}\,y$$

$$= \frac{N}{2\pi}\left[1 + \frac{1+2s(1-s)}{24Ns(1-s)} + o\left(\frac{1}{N}\right)\right]^2\int\limits_0^1\!\!\int \frac{\delta(\varphi-y+x)}{\sqrt{xy(1-x)(1-y)}}$$

$$\exp\left[-N\frac{(x-s)^2+(y-s)^2}{2s(1-s)} + N(1-2s)\frac{(x-s)^3+(y-s)^3}{3s^2(1-s)^2} - N(1-3s(1-s))\frac{(x-s)^4+(y-s)^4}{4s^3(1-s)^3} + \dots\right]\mathrm{d}\,x\,\mathrm{d}\,y. \tag{B7}$$

The previous expression suggests the introduction of the set of variables

$$\xi := \sqrt{N}(x-s), \tag{B8a}$$

$$\eta := \sqrt{N}(y-s), \tag{B8b}$$

in such a way that the integral becomes

$$\frac{\Gamma^2(N+1)}{\Gamma^2(Ns+{}^1\!/_2)\Gamma^2(N+{}^1\!/_2-Ns)}\int\limits_0^1\!\!\int \frac{\delta(\varphi-y+x)}{\sqrt{xy(1-x)(1-y)}}\,\mathrm{e}^{Ns\ln(xy)+N(1-s)\ln[(1-x)(1-y)]}\,\mathrm{d}\,x\,\mathrm{d}\,y$$

$$= \frac{1}{2\pi s(1-s)}\left[1 + \frac{1+2s(1-s)}{24Ns(1-s)} + o\left(\frac{1}{N}\right)\right]^2\int\limits_{-\infty}^{+\infty}\!\!\int \frac{\delta\left(\varphi - \frac{\eta}{\sqrt{N}} + \frac{\xi}{\sqrt{N}}\right)}{\sqrt{\left(1 - \frac{\xi}{(1-s)\sqrt{N}}\right)\left(1 - \frac{\eta}{(1-s)\sqrt{N}}\right)\left(1 + \frac{\xi}{s\sqrt{N}}\right)\left(1 + \frac{\eta}{s\sqrt{N}}\right)}}$$

$$\times \exp\left[-\frac{\xi^2+\eta^2}{2s(1-s)} + \frac{1-2s}{3s^2(1-s)^2}\frac{\xi^3+\eta^3}{\sqrt{N}} - \frac{1-3s(1-s)}{4(1-s)^3s^3}\frac{\xi^4+\eta^4}{N} + o\left(\frac{1}{N}\right)\right]\mathrm{d}\,\xi\,\mathrm{d}\,\eta. \tag{B9}$$

Introducing $\phi = \sqrt{N}\varphi$, the distribution in Eq. (52) is obtained through a series expansion for $N \gg 1$ and performing the Gaussian integrals.

Appendix C: On the minimum of asymptotically uncorrelated exchangeable variables

Let us consider a vector $\boldsymbol{\Phi}$ of $2N$ continuous variables, $\boldsymbol{\Phi} = (\Phi_1, \dots, \Phi_{2N})$, and let us assume that their joint probability distribution density is given by $p(\boldsymbol{\Phi})$. We assume that the $2N$ random variables are exchangeable,

i.e., such that $p(\Phi_1, \ldots, \Phi_{2N}) = p(\Phi_{\pi(1)}, \ldots, \Phi_{\pi(2N)})$ for any permutation $\pi \in \mathcal{S}_{2N}$ [28]. In the following, we will denote the expectation respect to the probability density $p(\boldsymbol{\Phi})$ by $\mathbb{E}(\bullet)$. Exchangeability implies

$$\mu := \mathbb{E}\left(\Phi_i\right), \tag{C1a}$$

$$\sigma^2 := \mathbb{E}\left[\left(\Phi_i - \mu\right)^2\right], \tag{C1b}$$

that we suppose to remain finite for $N \to +\infty$. We also make the assumption that the $2N$ components of the vector $\boldsymbol{\Phi}$ are weakly correlated, i.e., the covariance is given by

$$\frac{\rho}{N} := \mathbb{E}\left[\left(\Phi_i - \mu\right)\left(\Phi_j - \mu\right)\right], \tag{C1c}$$

in such a way that it vanishes as $O(1/N)$ when $N \to +\infty$, asymptotically recovering independence. Given a subset $\{\Phi_{l(i)}\}_{i=1,\ldots,K}$ of $1 \leq K \leq 2N$ different components of

$\boldsymbol{\Phi}$, it is easily seen that

$$0 \leq \mathbb{E}\left[\left(\sum_{i=1}^{K}\Phi_{l(i)} - K\mu\right)^2\right] = K\sigma^2 + K\frac{K-1}{N}\rho. \tag{C2}$$

For $K = 2N$, the relation above implies, for $N \to +\infty$, $\sigma^2 + 2\rho \geq 0$, whereas from $K = N$ we obtain $\sigma^2 + \rho > 0$ in the same limit. Let us now partition the $2N$ components of $\boldsymbol{\Phi}$ in two subsets with the same cardinality, for example the entries with even and odd labels, and consider

$$\epsilon_{\mathrm{o}} := \frac{1}{N}\sum_{i=1}^{N}\Phi_{2i-1}, \quad \epsilon_{\mathrm{e}} := \frac{1}{N}\sum_{i=1}^{N}\Phi_{2i}. \tag{C3}$$

We want to evaluate the mean value of $\epsilon := \min\{\epsilon_{\mathrm{o}}, \epsilon_{\mathrm{e}}\}$ for $N \gg 1$. The joint distribution of ϵ_{o} and ϵ_{e} is

$$\begin{aligned}
P(\epsilon_{\mathrm{o}}, \epsilon_{\mathrm{e}}) &= \mathbb{E}\left[\delta\left(\epsilon_{\mathrm{o}} - \frac{1}{N}\sum_{i=1}^{N}\Phi_{2i-1}\right)\delta\left(\epsilon_{\mathrm{e}} - \frac{1}{N}\sum_{i=1}^{N}\Phi_{2i}\right)\right] \\
&= \int\!\!\!\!\int_{-\infty}^{+\infty}\frac{\mathrm{d}\lambda_{\mathrm{e}}\,\mathrm{d}\lambda_{\mathrm{o}}}{4\pi^2}\mathbb{E}\left\{\exp\left[-i\lambda_{\mathrm{e}}\left(\epsilon_{\mathrm{e}} - \frac{1}{N}\sum_{k=1}^{N}\Phi_{2k}\right) - i\lambda_{\mathrm{o}}\left(\epsilon_{\mathrm{o}} - \frac{1}{N}\sum_{k=1}^{N}\Phi_{2k-1}\right)\right]\right\} \\
&= \int\!\!\!\!\int_{-\infty}^{+\infty}\mathrm{d}\lambda_{\mathrm{e}}\,\mathrm{d}\lambda_{\mathrm{o}}\frac{\exp\left[-i\lambda_{\mathrm{e}}(\epsilon_{\mathrm{e}} - \mu) - i\lambda_{\mathrm{o}}(\epsilon_{\mathrm{o}} - \mu) - \frac{\lambda_{\mathrm{e}}^2+\lambda_{\mathrm{o}}^2}{2N}\sigma^2 - \frac{(\lambda_{\mathrm{e}}+\lambda_{\mathrm{o}})^2}{2N}\rho\right]}{4\pi^2}\left[1 + o\left(\frac{1}{N}\right)\right]. \\
&= \frac{N}{2\pi\sigma\sqrt{\sigma^2 + 2\rho}}\exp\left[-\frac{N}{2\sigma^2\left(\sigma^2 + 2\rho\right)}\left(\epsilon_{\mathrm{e}} - \mu \quad \epsilon_{\mathrm{o}} - \mu\right)\begin{pmatrix}\sigma^2 + \rho & -\rho \\ -\rho & \sigma^2 + \rho\end{pmatrix}\begin{pmatrix}\epsilon_{\mathrm{e}} - \mu \\ \epsilon_{\mathrm{o}} - \mu\end{pmatrix}\right]\left[1 + o\left(\frac{1}{N}\right)\right], \tag{C4}
\end{aligned}$$

that is a bivariate Gaussian distribution function. The distribution $P(\epsilon)$ of the minimum $\epsilon := \min\{\epsilon_e, \epsilon_o\}$ is therefore given by

$$\begin{aligned}
P(\epsilon) &= -\frac{\partial}{\partial\epsilon}\int\!\!\!\!\int_{-\infty}^{+\infty}\theta(\epsilon_{\mathrm{e}} - \epsilon)\theta(\epsilon_{\mathrm{o}} - \epsilon)P(\epsilon_{\mathrm{e}}, \epsilon_{\mathrm{o}})\,\mathrm{d}\epsilon_{\mathrm{e}}\,\mathrm{d}\epsilon_{\mathrm{o}} \\
&= 2\int_{\epsilon}^{\infty}P(\hat{\epsilon}, \epsilon)\,\mathrm{d}\hat{\epsilon},
\end{aligned} \tag{C5}$$

that gives, in the limits of our approximations,

$$\begin{aligned}
P(\epsilon) = \sqrt{\frac{N}{2\pi}}\frac{\mathrm{e}^{-\frac{N}{2}\frac{(\epsilon-\mu)^2}{\sigma^2+\rho}}}{\sqrt{\sigma^2 + \rho}} \\
\times\left[1 - \mathrm{erf}\left(\sqrt{\frac{N}{2}}\frac{(\epsilon - \mu)\sigma}{\sqrt{(\sigma^2 + 2\rho)(\sigma^2 + \rho)}}\right)\right], \tag{C6}
\end{aligned}$$

with $\mathrm{erf}(x) := 2/\sqrt{\pi}\int_0^x \exp(-z^2)\,\mathrm{d}z$. In this form we see that the function $P(\epsilon)$ depends on $x = \sqrt{N}(\epsilon - \mu)$ and the distribution is a product of an even function and an odd function of x. In the spirit of the approximation, the domain $\mathscr{D}$ of x is substituted, for $N \gg 1$, with the entire real line, up to exponentially small corrections, being the probability distribution concentrated around μ. We immediately obtain

$$\int_{\mathscr{D}} P(\epsilon)\,\mathrm{d}\epsilon \simeq \int_{-\infty}^{\infty}\frac{\mathrm{e}^{-\frac{1}{2}\frac{x^2}{\sigma^2+\rho}}}{\sqrt{\sigma^2 + \rho}}\frac{\mathrm{d}x}{\sqrt{2\pi}} = 1. \tag{C7a}$$

$$N\int_{\mathscr{D}}(\epsilon - \mu)^2 P(\epsilon)\,\mathrm{d}\epsilon \simeq \int_{-\infty}^{\infty}\frac{\mathrm{e}^{-\frac{1}{2}\frac{x^2}{\sigma^2+\rho}}}{\sqrt{\sigma^2 + \rho}}\frac{\mathrm{d}x}{\sqrt{2\pi}} = \sigma^2 + \rho. \tag{C7b}$$

More interestingly

$$\sqrt{N} \int_{\mathscr{D}} (\epsilon - \mu) P(\epsilon)\, \mathrm{d}\,\epsilon \simeq$$

$$\simeq -\int_{-\infty}^{\infty} x \frac{\mathrm{e}^{-\frac{1}{2}\frac{x^2}{\sigma^2+\rho}}}{\sqrt{\sigma^2+\rho}} \mathrm{erf}\left(\sqrt{\frac{1}{2}} \frac{x\,\sigma}{\sqrt{(\sigma^2+2\rho)(\sigma^2+\rho)}}\right) \frac{\mathrm{d}\,x}{\sqrt{2\pi}}$$

$$= -\frac{\sigma}{\sqrt{\pi}}, \quad \text{(C7c)}$$

a result showing that, up to higher order terms, there is no influence of the weak correlation ρ on the expectation value of ϵ. This can be seen in a different way introducing the variables

$$X_- := \sqrt{N}\frac{\epsilon_{\mathrm{o}}-\epsilon_{\mathrm{e}}}{\sqrt{2}}, \quad X_+ := \sqrt{N}\frac{\epsilon_{\mathrm{o}}+\epsilon_{\mathrm{e}}-2\mu}{\sqrt{2}} \quad \text{(C8)}$$

in the distribution $P(\epsilon_{\mathrm{e}}, \epsilon_{\mathrm{o}})$, that gives the new distribution

$$P_X(X_-, X_+) = \frac{\exp\left[-\frac{1}{2(\sigma^2+2\rho)}X_+^2 - \frac{1}{2\sigma^2}X_-^2\right]}{2\pi\sqrt{\sigma^2(\sigma^2+2\rho)}}. \quad \text{(C9)}$$

Using the fact that, due to exchangeability, $P(\epsilon_{\mathrm{o}}, \epsilon_{\mathrm{e}}) = P(\epsilon_{\mathrm{e}}, \epsilon_{\mathrm{o}})$, we can replace

$$2(\epsilon_{\mathrm{e}}-\mu)\theta(\epsilon_{\mathrm{o}}-\epsilon_{\mathrm{e}}) \to (\epsilon_{\mathrm{e}}-\mu)\theta(\epsilon_{\mathrm{o}}-\epsilon_{\mathrm{e}}) + (\epsilon_{\mathrm{o}}-\mu)\theta(\epsilon_{\mathrm{e}}-\epsilon_{\mathrm{o}})$$

$$= \epsilon_{\mathrm{o}} - \mu - (\epsilon_{\mathrm{o}}-\epsilon_{\mathrm{e}})\,\theta\,(\epsilon_{\mathrm{o}}-\epsilon_{\mathrm{e}})$$

$$\to \sqrt{2}\frac{X_+ - X_-\theta(X_-)}{\sqrt{N}}. \quad \text{(C10)}$$

and therefore

$$\sqrt{N}\int_{-\infty}^{+\infty}(\epsilon-\mu)P(\epsilon)\,\mathrm{d}\,\epsilon =$$

$$= 2\sqrt{N}\iint_{-\infty}^{+\infty}(\epsilon_{\mathrm{e}}-\mu)\,\theta(\epsilon_{\mathrm{o}}-\epsilon_{\mathrm{e}})P(\epsilon_{\mathrm{e}},\epsilon_{\mathrm{o}})\,\mathrm{d}\,\epsilon_{\mathrm{e}}\,\mathrm{d}\,\epsilon_{\mathrm{o}}$$

$$= \sqrt{2}\iint_{-\infty}^{+\infty}[X_+ - X_-\theta(X_-)]\,P_X(X_-,X_+)\,\mathrm{d}\,X_+\,\mathrm{d}\,X_-$$

$$= -\frac{1}{\sqrt{\pi}\sigma}\int_0^{+\infty}X_-\exp\left[-\frac{1}{2\sigma^2}X_-^2\right]\mathrm{d}\,X_- = -\frac{\sigma}{\sqrt{\pi}}. \quad \text{(C11)}$$

[1] L. Lovász and D. Plummer, *Matching Theory*, AMS Chelsea Publishing Series (2009).

[2] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity* (Courier Dover Publications, 1998); H. W. Kuhn, Naval Res. Logistics Quart. **2**, 83 (1955); J. Munkres, Journal of the Society for Industrial and Applied Mathematics **5**, pp. 32 (1957); J. Edmonds, Can. J. Math. **17**, 449 (1965); R. Jonker and A. Volgenant, Computing **38**, 325 (1987); D. Jungnickel and T. Schade, *Graphs, networks and algorithms* (Springer, 2005).

[3] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford Graduate Texts (OUP Oxford, 2009).

[4] M. Mézard and G. Parisi, J. Phys. Lett. **46**, 771 (1985).

[5] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond*, Lecture Notes in Physics Series (World Scientific Publishing Company, Inc., 1987).

[6] M. Mézard and G. Parisi, J. Phys. **49**, 2019 (1988).

[7] G. Sicuro, *The Euclidean matching problem*, Springer theses (Springer, 2017).

[8] S. Kirkpatrick, C. Gelatt, and M. Vecchi, Science **220**, 671 (1983).

[9] H. Orland, J. Phys. Lett. **46**, 763 (1985).

[10] M. Mézard and G. Parisi, J. Phys. **48**, 1451 (1987); G. Parisi and M. Ratiéville, Eur. Phys. J. B **29**, 457 (2002); S. Caracciolo, M. P. D'Achille, E. M. Malatesta, and G. Sicuro, Phys. Rev. E **95**, 052129 (2017); C. Lucibello, G. Parisi, and G. Sicuro, Phys. Rev. E **95**, 012302 (2017).

[11] S. Caracciolo and G. Sicuro, Phys. Rev. Lett. **115**, 230601 (2015).

[12] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, Phys. Rev. E **90**, 012118 (2014).

[13] C. Villani, *Optimal transport: old and new* (Springer, 2008).

[14] L. Ambrosio, F. Stra, and D. Trevisan, (2016), arXiv:1611.04960 [math].

[15] S. Bobkov and M. Ledoux, preprint (2016), *One-dimensional empirical measures, order statistics and Kantorovich transport distances*.

[16] R. J. McCann, Proc. R. Soc. A: Math., Phys. Eng. Sci. **455**, 1341 (1999).

[17] E. Boniolo, S. Caracciolo, and A. Sportiello, J. Stat. Mech. **2014**, P11023 (2014).

[18] S. Caracciolo and G. Sicuro, Phys. Rev. E **90**, 042112 (2014).

[19] S. Caracciolo and G. Sicuro, Phys. Rev. E **91**, 062125 (2015).

[20] H. Orland and A. Zee, Nucl. Phys. B **620**, 456 (2002).

[21] S. K. Nechaev, A. N. Sobolevski, and O. V. Valba, Phys. Rev. E **87**, 012102 (2013).

[22] M. D. Donsker, Ann. Math. Statist. **23**, 277 (1952); R. Dudley, *Uniform Central Limit Theorems*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, 1999).

[23] J. Komlós, P. Major, and G. Tusnády, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **32**, 111 (1975).

[24] M. Csörgő and P. Révész, *Strong approximations in probability and statistics*, Probability and mathematical statistics (Academic Press, 1981).

[25] This result has been obtained, for example, in Ref. [17]. Observe, however, that in Ref. [17] one set of points is

supposed to be fixed and, for this reason, the variance of the distribution is one half of the variance of the leading term in Eq. (52).

[26] It is amusing to remark that if we expand $6\overline{\tau^2}$ around $p = -\infty$ we get a series with integer coefficients

$$6\overline{\tau^2} = \sum_{n \geq 0} \frac{a_n}{p^n}, \quad a_n = \begin{cases} 1 & \text{for } n = 0, \\ 3^{n-2}(n+2) & \text{for } n > 0. \end{cases}$$

The sequence $(a_n)_{n \in \mathbb{N}}$ can have a combinatorial interpretation. It is indeed the number of spanning trees in a $(n-2)$-book graph [29].

[27] R. Pyke, J. R. Stat. Soc. Series B **27**, 395 (1965).

[28] D. J. Aldous, "Exchangeability and related topics," in *École d'Été de Probabilités de Saint-Flour XIII — 1983*, edited by P. L. Hennequin (Springer Berlin Heidelberg, Berlin, Heidelberg, 1985) pp. 1–198.

[29] *The On-Line Encyclopedia of Integer Sequences*, sequence A006234; T. Došlić, J. Math. Chemistry **51**, 1599 (2013).