On the one dimensional Euclidean matching problem: exact solutions, correlation functions and universality

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

Gabriele Sicuro
Dipartimento di Fisica "E. Fermi", University of Pisa and INFN, Largo Bruno Pontecorvo, 3, I-56127 Pisa, Italy†

(Dated: August 16, 2014)

We discuss the equivalence relation between the Euclidean bipartite matching problem on the line and on the circumference and the Brownian bridge process on the same domains. The equivalence allows us to compute the correlation function and the optimal cost of the original combinatoric problem in the thermodynamic limit; moreover, we solve also the minimax problem on the line and on the circumference. The properties of the average cost and correlation functions are discussed.

I. INTRODUCTION

Let us consider a bipartite complete graph with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, $\mathfrak{K}_{N,N} = \mathrm{Graph}(\mathcal{V}, \mathcal{E})$, whose vertex set can be partitioned in two disjoint subsets of the same cardinality, $\mathcal{V} = \mathcal{R} \cup \mathcal{B}$, $|\mathcal{R}| = |\mathcal{B}| = N \in \mathbb{N}$, $\mathcal{R} \coloneqq \{r_1, \dots, r_N\}$, $\mathcal{B} \coloneqq \{b_1, \dots, b_N\}$. Let us also introduce a weight function $w\colon \mathcal{E} \to \mathbb{R}$, $(r_i b_j) \in \mathcal{E} \mapsto w_{ij}$. In the weighted bipartite matching problem, we are interested in the permutation π of N elements such that a certain cost function

$$E[\pi] \coloneqq f\left(w_{1\pi(1)}, \dots, w_{N\pi(N)}\right), \quad f\colon \mathbb{R}^N \to \mathbb{R}^+, \tag{1}$$

is minimized. The most common form adopted in the literature for the function f is simply $f(x_1, \dots, x_N) \coloneqq \sum_{i=1}^N |x_i|$: in this case, the assignment problem can be solved in polynomial time using the Khun–Munkres algorithm [1, 2] that, in the Edmonds and Karp's version [3], has $O(N^3)$ running time.

As a variation of the problem, sometimes random weights $\{w_{ij}\}$ are considered: in this case the average properties of the solution are of great interest. In the hypothesis of independently and identically distributed weights, the problem was studied using arguments borrowed both from probability theory [4] and from the theory of disordered systems [5]. Finally, if the vertices $\mathcal{V}$ of the graph are associated to points randomly generated in the hypercube $\Omega = [0,1]^d$ in d dimensions and w_{ij} is a function of the Euclidean distance between the r_i vertex and the b_j vertex, the problem is called *Euclidean bipartite matching problem* [6–8].

In the present paper we will consider the so-called *grid-Poisson matching problem* in one dimension both with open boundary conditions (OBC) and with periodic boundary conditions (PBC). The set of vertices $\mathcal{R}$ is associated to a set of fixed points on the interval $\Omega \coloneqq [0,1]$ and in particular $r_i \mapsto x_i \equiv \frac{2i-1}{2N}$, $i = 1, \dots, N$, whilst the set of vertices $\mathcal{B}$ is associated to a set of N points, $\{y_i\}_{i=1,\dots,N}$, randomly generated in the interval, such that $b_i \mapsto y_i$. We will suppose the $\mathcal{B}$-vertices indexed in such a way that $i \le j \Rightarrow y_i \le y_j$. Finally, we will consider the following cost functional

$$E_p[\pi] = \sqrt[p]{\frac{1}{N}\sum_{i=1}^N \left[w\left(|x_i - y_{\pi(i)}|\right)\right]^p}, \quad p \in (1, +\infty), \tag{2}$$

in which the function $w\colon [0,1] \to [0,1]$ is defined as below:

$$w(x) = \begin{cases} x & \text{for OBC,} \\ x\,\theta\left(\frac{1}{2} - x\right) + (1-x)\,\theta\left(x - \frac{1}{2}\right) & \text{for PBC,} \end{cases} \tag{3}$$

where $\theta(x)$ is the Heaviside function. In the following,

$$\epsilon_p \coloneqq \sqrt{N} \min_\pi E_p[\pi], \tag{4}$$

and

$$\mathfrak{e}_p \coloneqq \sqrt{N^p} \min_\pi \left[(E_p[\pi])^p\right]. \tag{5}$$

We will show that, for the cost functional (2), a solution of the problem can be obtained in the continuum limit, $N \to \infty$, not only for $p = 2$ (as already shown in Ref. [7]) but also in the $p \to \infty$ limit using well known properties of the Brownian bridge process. In fact, in the limit $p \to \infty$,

$$\epsilon_p \xrightarrow{p\to\infty} \sqrt{N} \min_\pi \max_i w\left(|x_i - y_{\pi(i)}|\right), \tag{6}$$

i.e., the problem reduces to the minimax grid-Poisson matching problem in one dimension. The problem was studied by Leighton and Shor [9] for $d = 2$ and Shor and Yukich [10] for $d \ge 3$. The minimax problem is related to a lot of different computational problems and the evaluation of the scaling of its cost gives directly useful informations on the computational cost of other algorithms (for a discussion of the related problems in $d = 2$ see for example Ref. [9]).

*Electronic address: sergio.caracciolo@mi.infn.it
†Electronic address: gabriele.sicuro@for.unipi.it

II. OPTIMAL COST AND CORRELATION FUNCTION ON THE INTERVAL

The solution of the grid–Poisson matching problem in one dimension for OBC is easily found by simple arguments [7] for $p > 1$ and cost functional (2): in this case, in fact, the optimal matching is always *ordered*, i.e. $\pi(i) = i$, in the hypothesis above that $i < j \Rightarrow y_i \leq y_j$. Remarkably, the solution is very different if $p \in (0, 1)$, as discussed in [7, 11]: indeed, in that case the solution is expressed in terms of non intersecting arcs joining the matched points on the segment. Note that, due to the monotony of the function $x \mapsto x^p$, the optimal solution for the cost E_p is also optimal for the cost $(E_p)^p$. The probability density distribution for the position of the i-th $\mathcal{B}$-point is:

$$\Pr\left(y_i \in \mathrm{d}\, y\right) = \binom{N}{i} y^i (1-y)^{N-i} \frac{i}{y}\, \mathrm{d}\, y, \tag{7}$$

where we used the short-hand notation $x \in \mathrm{d}\, z \Leftrightarrow x \in [z, z + \mathrm{d}\, z]$. In the $N \to \infty$ limit, a non trivial result is obtained introducing the variable $m(y)$

$$m(y) \coloneqq \sqrt{N} M(y), \quad M(y) \coloneqq \frac{i}{N} - y \tag{8}$$

expressing the rescaled (signed) distance between a $\mathcal{B}$-point in $[y, y + \mathrm{d}\, y]$ and its corresponding $\mathcal{R}$-point in the optimal matching. We finally obtain a distribution for the variable $m(y)$ depending on the position on the interval $y \in [0, 1]$:

$$\Pr\left(m(y) \in \mathrm{d}\, m\right) = \frac{\mathrm{e}^{-\frac{m^2}{2y(1-y)}}}{\sqrt{2\pi y(1-y)}}\, \mathrm{d}\, m. \tag{9}$$

The distribution (9) is the one of a Brownian bridge on the interval $[0, 1]$, a continuous time stochastic process defined as

$$\mathsf{B}(t) \coloneqq \mathsf{W}(t) - t\mathsf{W}(1), \quad t \in [0, 1], \tag{10}$$

where $\mathsf{W}(t)$ is a Wiener process. The joint distribution of the process can be derived similarly (see Appendix). In particular, the covariance matrix for the 2-points joint distribution has the form, for $s, t \in [0, 1]$ (see eq. (A6)),

$$\Sigma_2 = \begin{pmatrix} 2\phi(s) & \phi(s) + \phi(t) - \phi(t-s) \\ \phi(s) + \phi(t) - \phi(t-s) & 2\phi(t) \end{pmatrix}, \tag{11}$$

where we introduced the function

$$\phi(x) \coloneqq |x| \frac{1 - |x|}{2}. \tag{12}$$

Averaging over the positions s, t and fixing the distance $\tau \coloneqq |t - s|$, we have

$$\bar{\Sigma}_2(\tau) = \begin{pmatrix} c & c - \phi(\tau) \\ c - \phi(\tau) & c \end{pmatrix}, \quad c = \frac{1}{6}. \tag{13}$$

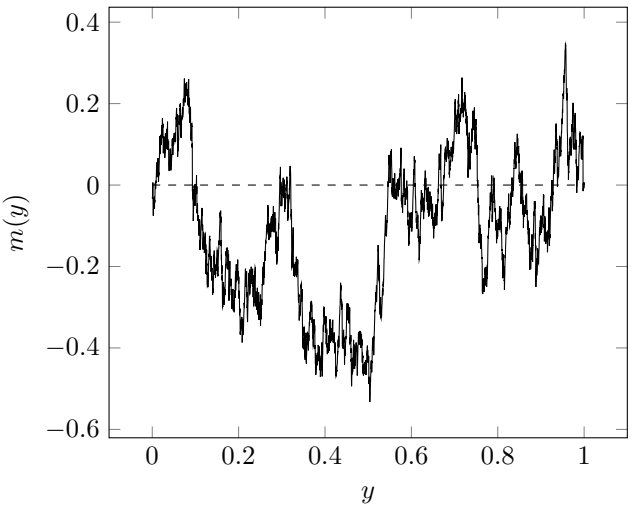

Figure 1: Plot of $m(y_i) = \sqrt{N}\left(\frac{i}{N} - y_i\right)$ for a certain realisation of the problem with $N = 4096$.

The Euclidean matching problem on the interval $[0, 1]$ with open boundary conditions and cost functional (2) is therefore related to the Brownian bridge in the $N \to \infty$ limit. By using this correspondence, the correlation function $\forall p > 1$ is computed as

$$\overline{\langle m(t) m(t+\tau) \rangle} = \frac{1}{6} - \phi(\tau). \tag{14}$$

where the average $\bar{\bullet}$ is intended on the position t, whilst we denoted by $\langle \bullet \rangle$ the average over different realisations of the problem. This theoretical prediction was confirmed numerically. Introducing the normalised variable

$$\sigma(t) = \frac{m(t)}{|m(t)|} = \mathrm{sign}(m(t)), \tag{15}$$

Boniolo et al. [7] computed also the correlation function for this quantity, finding

$$\langle \sigma(s)\sigma(t) \rangle = \frac{2}{\pi} \arctan \sqrt{\frac{\min(s,t)(1 - \max(s,t))}{|t-s|}}$$
$$\Rightarrow \int_0^{1-t} \langle \sigma(s)\sigma(s+t) \rangle\, \mathrm{d}\, s = \frac{1 - \sqrt{t}}{1 + \sqrt{t}}. \tag{16}$$

Both formulas were confirmed numerically. Note that all the results above holds $\forall p > 1$ in the case of open boundary conditions.

Let us now compute the average cost of the matching. From eq. (9) we obtained that

$$\langle \mathfrak{e}_p \rangle \xrightarrow{N \to \infty} \int_0^1 \langle |\mathsf{B}(t)|^p \rangle\, \mathrm{d}\, t = \frac{1}{2^{\frac{p}{2}}(p+1)} \Gamma\left(\frac{p}{2} + 1\right). \tag{17}$$

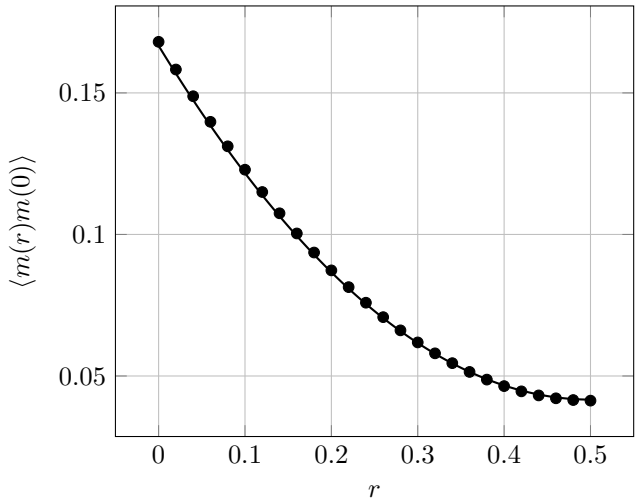

Figure 2: Correlation function $\forall p > 1$ in the OBC case for $N = 1000$ obtained averaging over 10^4 realisations (for clarity, not all data are represented).

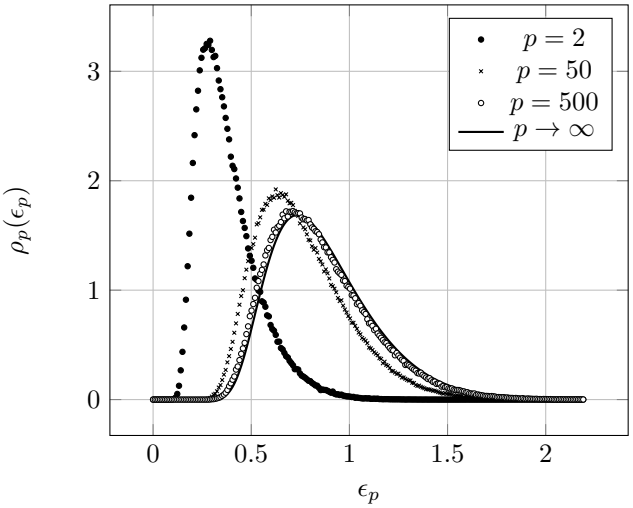

Figure 3: Distribution density of the rescaled cost for different values of p computed with $5 \cdot 10^5$ iterations and $N = 5000$ for the OBC case.

Moreover, the optimal cost ϵ_p in the $N \to \infty$ limit can be written as

$$\langle \epsilon_p \rangle \xrightarrow{N\to\infty} \left\langle \left(\int_0^1 |\mathsf{B}(t)|^p \, \mathrm{d}\, t \right)^{\frac{1}{p}} \right\rangle . \tag{18}$$

Although the previous expression is difficult to evaluate exactly for finite p (see for example Ref. [12] for additional information about the distribution of $\mathsf{X}_p \coloneqq \int_0^1 |\mathsf{B}(t)|^p \, \mathrm{d}\, t$), the calculation can be easily performed in the relevant limit $p \to \infty$, being

$$\left\langle \left(\int_0^1 |\mathsf{B}(t)|^p \, \mathrm{d}\, t \right)^{\frac{1}{p}} \right\rangle \xrightarrow{p\to\infty} \left\langle \sup_{t\in[0,1]} |\mathsf{B}(t)| \right\rangle . \tag{19}$$

The distribution of the supremum of the absolute value of the Brownian bridge is the well known Kolmogorov distribution [13]

$$\Pr\left(\sup_{t\in[0,1]} |\mathsf{B}(t)| < u \right) = \sum_{k=-\infty}^{+\infty} (-1)^k \, \mathrm{e}^{-2k^2u^2} \tag{20}$$

and therefore

$$\langle \epsilon_p \rangle \xrightarrow{N\to\infty} \sqrt{\frac{\pi}{2}} \ln 2. \tag{21}$$

In figure 3 we plotted $\rho_p(u) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\, u} \left[\Pr\left(\epsilon_p \leq u \right) \right]$ for different values of p: observe that ρ_p approaches the Kolmogorov distribution in the large p limit.

Finally, observe also that the variance of $\mathfrak{e}_p$

$$\left\langle \left(\mathfrak{e}_p - \langle \mathfrak{e}_p \rangle \right)^2 \right\rangle = \frac{(1+p)\Gamma(p+2) - (2p+1)\Gamma^2\left(\frac{p}{2}+1\right)}{2^p (p+1)^2 (2p+1)} \tag{22}$$

increases with p and that

$$\frac{\left\langle \left(\mathfrak{e}_p - \langle \mathfrak{e}_p \rangle \right)^2 \right\rangle}{\langle \mathfrak{e}_p \rangle^2} = \frac{(p+1)\Gamma(p+2)}{(2p+1)\Gamma\left(\frac{p}{2}+1\right)^2} - 1 \xrightarrow{p\to\infty} +\infty. \tag{23}$$

From a numerical point of view, this means that a computation of $\langle \mathfrak{e}_p \rangle$ requires a larger amount of iterations as p increases and fluctuations around the mean value become extremely large in the $p \to \infty$ limit. On the other hand, fluctuations of the optimal cost ϵ_p around its mean value $\langle \epsilon_p \rangle$ for $p \to \infty$ remain finite

$$\left\langle \left(\epsilon_p - \langle \epsilon_p \rangle \right)^2 \right\rangle \xrightarrow{p\to\infty} \frac{\pi^2}{12} - \frac{\pi}{2} \ln^2 2. \tag{24}$$

This fact allows to perform a precise computation of $\langle \epsilon_p \rangle$ for large p.

III. OPTIMAL COST AND CORRELATION FUNCTION ON THE CIRCUMFERENCE

Let us now consider the case of periodic boundary conditions. As discussed in Ref. [7], the solution for both the cost E_p and the cost $(E_p)^p$, with $p \in (1, +\infty)$, is again ordered; however, in this case the mapping is $i \mapsto \pi(i) = i + \lambda \mod N$ for a certain $\lambda \in \{0, 1, \ldots, N-1\}$. In the continuum limit, the solution is a generalised Brownian bridge, $m_p(t) = \mathsf{B}(t) + \lambda_p$, $t \in [0,1]$, for a certain constant $\lambda_p \in \mathbb{R}$ depending on p. The constant λ_p can be found by optimality condition on the cost functional (2):

$$\frac{\partial}{\partial \lambda_p} \left(\int_0^1 |\mathsf{B}(t) + \lambda_p|^p \, \mathrm{d}\, t \right)^{\frac{1}{p}} = 0. \tag{25}$$

The previous equation can be solved only for $p = 2$ and $p \to +\infty$. For $p = 2$, $\lambda_2 = -\int_0^1 \mathsf{B}(\tau) \, \mathrm{d}\, \tau$; therefore,

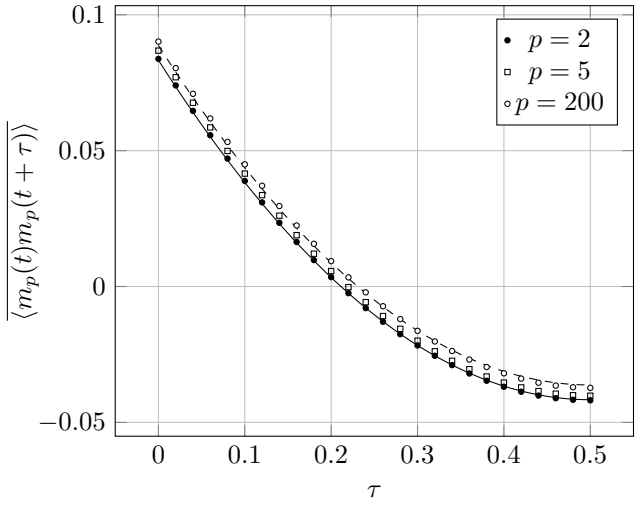

Figure 4: Correlation function in the PBC case for $N = 1000$ obtained averaging over 10^4 realisations (for clarity, not all data are represented). The continuous line is the theoretical prediction for $p = 2$; the dashed line corresponds to the $p \to \infty$ case.

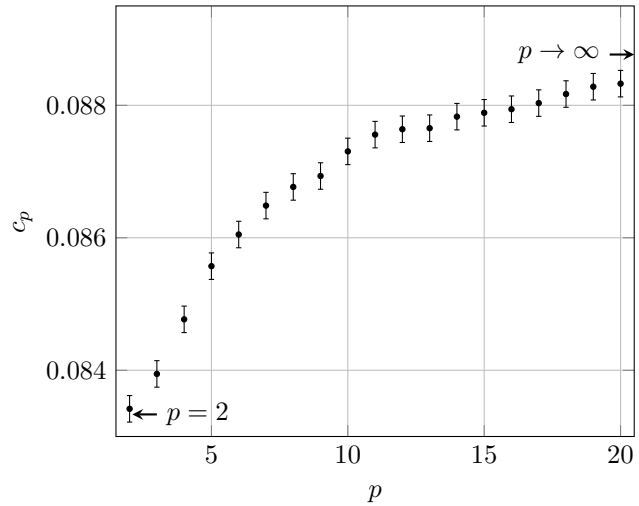

Figure 5: Value of $c_p \coloneqq \overline{\langle m_p^2(t)\rangle}$ in the PBC case for $N = 1000$ obtained from the fit of the correlation function, averaging over 5000 realisations, for different values of p.

$m_2(t) = \mathsf{B}(t) - \int_0^1 \mathsf{B}(t)\,\mathrm{d}\,t$ and

$$\overline{\langle m_2(t)m_2(t+\tau)\rangle} = \frac{1}{12} - \phi(\tau). \tag{26}$$

For $p \to \infty$, $\left(\int_0^1 |\mathsf{B}(t) + \lambda_p|^p\,\mathrm{d}\,t\right)^{\frac{1}{p}} \xrightarrow{p\to\infty} \sup_{t\in[0,1]} |\mathsf{B}(t) + \lambda_\infty|$ and therefore the optimality condition becomes

$$\frac{\partial}{\partial\lambda}\left(\sup_{t\in[0,1]} |\mathsf{B}(t) + \lambda|\right) = 0$$
$$\Rightarrow \lambda_\infty = -\frac{\sup_{t\in[0,1]} \mathsf{B}(t) + \inf_{t\in[0,1]} \mathsf{B}(t)}{2}. \tag{27}$$

Indeed, we have that

$$\begin{aligned}&\sup_t |\mathsf{B}(t) + \lambda| = \\ &= |\sup_t \mathsf{B}(t) + \lambda|\theta(|\sup_t \mathsf{B}(t) + \lambda| - |\inf_t \mathsf{B}(t) + \lambda|) \\ &+ |\inf_t \mathsf{B}(t) + \lambda|\theta(|\inf_t \mathsf{B}(t) + \lambda| - |\sup_t \mathsf{B}(t) + \lambda|)\end{aligned} \tag{28}$$

from which the eq. (27) is derived. We have therefore, for $t \in [0, 1]$,

$$m_\infty(t) = \mathsf{B}(t) - \frac{\sup_{s\in[0,1]} \mathsf{B}(s) + \inf_{s\in[0,1]} \mathsf{B}(s)}{2}. \tag{29}$$

The correlation function can be directly found using the known joint distributions for the Brownian bridge and its sup, eq. (B2), and for the sup and inf of a Brownian bridge, eq. (B1). After some calculations we obtain

$$\overline{\langle m_\infty(t)m_\infty(t+\tau)\rangle} = \frac{2 - \zeta(2)}{4} - \phi(\tau), \tag{30}$$

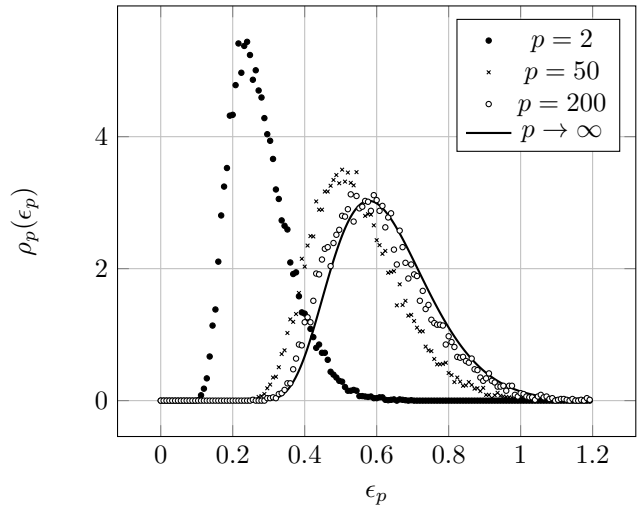

Figure 6: Density distribution of the rescaled cost for different values of p computed with $2 \cdot 10^4$ iterations and $N = 1000$ for the PBC case.

where $\zeta(2) \coloneqq \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. The value $\overline{\langle m_\infty^2(t)\rangle} = \frac{2-\zeta(2)}{4} = 0.0887665\ldots$ is very close to the value obtained for $p = 2$, $\overline{\langle m_2^2(t)\rangle} = \frac{1}{12} = 0.08\bar{3}$. In figure 5 we plot the values of $\overline{\langle m_p^2(t)\rangle}$ as function of p. Note, finally, that due to the fact that we have imposed PBC, $\overline{\langle\bullet\rangle} \equiv \langle\bullet\rangle$ holds in all the previous formulas.

Let us now introduce the normalised transport field

$$\sigma_p(t) \coloneqq \frac{m_p(t)}{|m_p(t)|} = \mathrm{sign}(m_p(t)). \tag{31}$$

The correlation function $\langle\sigma_p(s)\sigma_p(t)\rangle$ can be computed from the covariance matrix observing that the process is still Gaussian. The correlation function is found in the

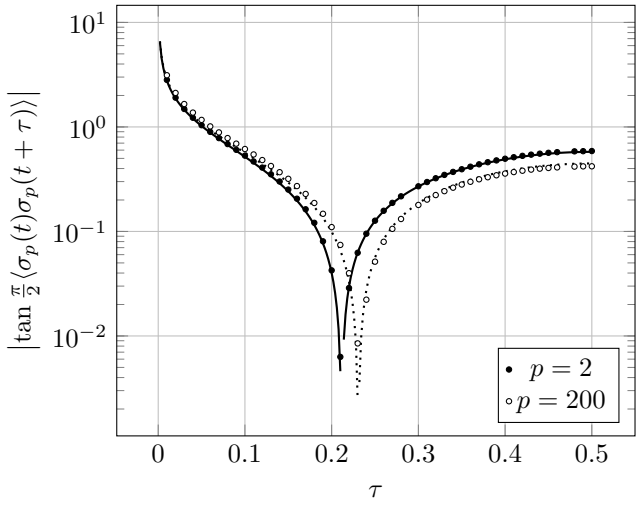

Figure 7: Plot for the normalised map correlation function (32) obtained with 10^4 iterations and $N = 1000$ for the PBC case. The continuous line is the theoretical prediction for $p = 2$, whilst the dotted line is obtained from the theoretical prediction for $p \to \infty$ The pole is at $\tau(p) = \frac{1}{2} - \frac{\sqrt{1-8c_p}}{2}$.

form

$$\langle\sigma_p(s)\sigma_p(t)\rangle = \frac{2}{\pi}\arctan\frac{c_p - \phi(t-s)}{\sqrt{\phi(t-s)\,(2c_p - \phi(t-s))}} \tag{32}$$

where $c_2 = \frac{1}{12}$ for $p = 2$ and $c_\infty = \frac{2-\zeta(2)}{4}$.

The optimal cost in the $p \to \infty$ limit can be evaluated as the average spread of the Brownian bridge. Denoting

$$\xi \coloneqq \sup_{s\in[0,1]} \mathsf{B}(s) - \inf_{s\in[0,1]} \mathsf{B}(s), \tag{33}$$

the distribution of the spread ξ is given by [13]

$$\Pr(\xi < u) = \theta_3\left(\mathrm{e}^{-2u^2}\right) + u\frac{\mathrm{d}}{\mathrm{d}\,u}\theta_3\left(\mathrm{e}^{-2u^2}\right), \tag{34}$$

where

$$\theta_3(x) \equiv \vartheta_3(0,x), \quad \vartheta_3(z,q) \coloneqq 1 + 2\sum_{n=1}^{\infty} q^{n^2}\cos(2nz) \tag{35}$$

is the third Jacobi theta function. From eq. (34) the distribution of the optimal cost in the $p \to \infty$ limit is easily derived. Moreover,

$$\begin{aligned}\langle\epsilon_p\rangle &\xrightarrow[p\to\infty]{N\to\infty} \langle\epsilon_\infty\rangle = \frac{\langle\xi\rangle}{2} = \\ &= \frac{1}{2}\left(\left\langle\sup_{s\in[0,1]}\mathsf{B}(s)\right\rangle - \left\langle\inf_{s\in[0,1]}\mathsf{B}(s)\right\rangle\right) = \frac{1}{2}\sqrt{\frac{\pi}{2}}\end{aligned} \tag{36}$$

with corresponding variance

$$\left\langle\left(\epsilon_p - \langle\epsilon_p\rangle\right)^2\right\rangle \xrightarrow{p\to\infty} \frac{\pi^2 - 3\pi}{24}. \tag{37}$$

IV. GENERAL SOLUTION: A CONTINUUM APPROACH

In the present section we will justify and generalise the previous results looking at a continuum version of the problem, the so called *Monge–Kantorovič problem*.

Let us consider the interval $\Omega \coloneqq [0, L] \subset \mathbb{R}$, $L \in \mathbb{R}^+$ and let us suppose also that two different measures are given on Ω, i.e., the uniform (Lebesgue) measure

$$\mathrm{d}\,m(x) \coloneqq \frac{1}{L}\,\mathrm{d}\,x, \tag{38}$$

and a non uniform measure $\mathrm{d}\,n(x)$ with measure density $\nu(x)$,

$$\begin{aligned}\mathrm{d}\,n(x) \coloneqq \nu(x)\,\mathrm{d}\,x = \tfrac{\mathrm{d}\,x}{L} + \\ \mathrm{d}\,x\sum_{k=1}^{\infty}\left(\nu_1(k)\cos\tfrac{2\pi kx}{L} + \nu_2(k)\sin\tfrac{2\pi kx}{L}\right).\end{aligned} \tag{39}$$

We ask for the optimal map $\mu(x)\colon \Omega \to \Omega$ such that the following transport condition is satisfied

$$\frac{1}{L}\int_A \mathrm{d}\,x = \int_{\mu^{-1}(A)} \mathrm{d}\,n(x) \quad \forall A \subset \Omega \text{ measurable.} \tag{40}$$

and μ minimises the following functional

$$\mathfrak{E}_p[\mu] \coloneqq \int_0^L |x - \mu(x)|^p\,\mathrm{d}\,n(x), \quad p \in \mathbb{R}^+. \tag{41}$$

It can be proved [14] that for $p > 1$ eq. (40) can be rewritten as a change-of-variable formula:

$$L\,\mathrm{d}\,n(x) = \mathrm{d}\,\mu(x). \tag{42}$$

We will restrict therefore to the case $p > 1$. Imposing PBC, that is $\mu_p(0) = \mu_p(L) - L$, the solution of (42) determines the optimal map up to a constant $\mu_p(0)$ as

$$\mu_p(x) = x + \mu_p(0) + \Phi(x). \tag{43}$$

In the previous equation we have introduced $\Phi(x)$

$$\begin{aligned}\Phi(x) \coloneqq \textstyle\sum_{k=1}^{\infty}\nu_1(k)\frac{L^2\sin\left(\frac{k\pi x}{L}\right)}{\pi k}\cos\left(\frac{k\pi x}{L}\right) \\ + \textstyle\sum_{k=1}^{\infty}\nu_2(k)\frac{L^2}{\pi k}\sin^2\left(\frac{k\pi x}{L}\right).\end{aligned} \tag{44}$$

Note that $\Phi(0) = \Phi(L) = 0$. The value of $\mu_p(0)$ must be determined requiring that the functional (41) is minimum: we have that

$$\begin{aligned}&\tfrac{\partial}{\partial\mu_p(0)}\textstyle\int_0^L |\mu_p(0) + \Phi(x)|^p\,\mathrm{d}\,n(x) = \\ &= p\textstyle\int_0^L \operatorname{sign}(\mu_p(0) + \Phi(x))\,|\mu_p(0) + \Phi(x)|^{p-1}\,\mathrm{d}\,n(x) = 0.\end{aligned} \tag{45}$$

If instead OBC are considered, then $\mu_p(0) = 0$ and the solution is obtained explicitly $\forall p > 1$.

Let us now suppose that $L \equiv N \in \mathbb{N}$ and that the measure $n(\mathrm{d}\,x)$ is obtained as a limit measure of a random atomic measure of the form

$$\begin{aligned}\mathrm{d}\,n_N(x) &\coloneqq \tfrac{\mathrm{d}\,x}{N}\sum_{i=1}^{N}\delta\left(x-\eta_i\right)\\ &= \mathrm{d}\left(\tfrac{1}{N}\sum_{i=1}^{N}\theta\left(x-\eta_i\right)\right),\end{aligned} \quad (46)$$

where $\{\eta_i\}_{i=1,\dots,N}$ is a set of N points uniformly randomly distributed in Ω. The previous measure can be written as

$$\begin{aligned}n_N(x) = \sum_{k=1}^{\infty}\sqrt{\tfrac{2}{N}}\tfrac{Z_k}{\pi k}\sin\tfrac{\pi k x}{N} + \tfrac{x}{N}\\ + \tfrac{1}{N}\sum_{i=1}^{N}\eta_i - 1.\end{aligned} \quad (47)$$

where we have introduced

$$\begin{aligned}Z_k \equiv Z_k(x) &\coloneqq \frac{1}{N}\sum_{i=1}^{N} z_i(x),\\ z_i(x) &\coloneqq -\sqrt{2N}\cos\left(k\pi\frac{2\eta_i + x}{N}\right).\end{aligned} \quad (48)$$

Observe now that $Z_k(x)$ is a sum of independent identically distributed random variables. From the central limit theorem, we have that $Z_k(x)$ is normally distributed as

$$Z_k \sim \mathcal{N}(0,1) \quad \forall k \in \mathbb{N}. \quad (49)$$

Remarkably the previous distribution does not depend on x. Moreover, the Z_k and Z_l are independent random variables for $k \neq l$, being Gaussian distributed and $\langle Z_l Z_k\rangle = 0$, where the average $\langle\bullet\rangle$ is intended over the possible values $\{\eta_i\}_i$. In eq. (47) the Karhunen–Loève expansion for the Brownian bridge [15] on the interval $[0, N]$ appears:

$$\mathsf{B}_N(x) \coloneqq \sum_{k=1}^{\infty}\sqrt{2N}\frac{Z_k}{\pi k}\sin\frac{\pi k x}{N}, \quad Z_k \sim \mathcal{N}(0,1) \quad \forall k \in \mathbb{N}. \quad (50)$$

It follows that $n_N(x)$ can be written for large N, up to irrelevant additive constants, as

$$n_N(x) \simeq \frac{1}{N}\mathsf{B}_N(x) + \frac{x}{N} \quad (51)$$

and therefore we cannot associate a density measure to it, due to the fact that $n_N(x)$ is not differentiable.

However the solution of the matching problem in the continuum can be still obtained directly from eq. (39); considering PBC, then

$$\mu_p(x) = \mu_p(0) + x + \mathsf{B}_N(x), \quad x \in [0, N]. \quad (52)$$

Denoting by

$$M_p(x) \coloneqq \mu_p(x) - x, \quad (53)$$

it follows that $\forall p > 1$

$$\langle M_p(x) M_p(y)\rangle = c_p(N) - N\phi\left(\frac{x-y}{N}\right). \quad (54)$$

where $c_p(N)$ is a constant depending on N and p. Adopting the notation

$$F(N) \sim G(N) \Leftrightarrow 0 < \lim_{N\to\infty}\frac{F(N)}{G(N)} < +\infty, \quad (55)$$

for two positive real functions F and G depending on N, note that

$$c_p(N) \sim c_p N \quad (56)$$

for some positive constant c_p depending on p. Indeed, in the discrete case, from the fact that the solution must be ordered, in the large N limit it can be easily seen that

$$\min_{\mu}\mathfrak{E}_p[\mu] \sim N^{\frac{p}{2}}, \quad (57)$$

where $\mathfrak{E}_p$ is the cost functional (41) in which the measure (46) is adopted. Therefore, $\mu_p(0) = M_p(0) \sim M_p(x) \sim \sqrt{N}$. Moreover, note also that $\overline{\langle M^2(t)\rangle} < N^2 \Rightarrow c_p(0) = 0$. For $p = 2$, eq. (45) becomes

$$\begin{aligned}\mu_2(0) &= -\tfrac{1}{N}\int_0^N \mathsf{B}_N(x)\circ\mathrm{d}\,\mathsf{B}_N(x) - \tfrac{1}{N}\int_0^N \mathsf{B}_N(x)\,\mathrm{d}\,x\\ &= -\tfrac{1}{N}\int_0^N \mathsf{B}_N(x)\,\mathrm{d}\,x\end{aligned} \quad (58)$$

where the first integral is intended in the Stratonovič sense. The result is in agreement with the one presented in Section III. If we consider the transport cost functional

$$E_p[\tilde{\mu}] \coloneqq \sqrt[p]{\int_0^N |x - \tilde{\mu}(x)|^p\,\mathrm{d}\,n(x)} \quad (59)$$

the matching problem has the same solution obtained for the cost (41) for all finite values of p, $\mu_p(x) \equiv \tilde{\mu}_p(x) \Rightarrow \mu_p(0) \equiv \tilde{\mu}_p(0)$, due to the fact that the function $f(x) = \sqrt[p]{x}$ is monotone. However, for the functional cost (59), in the $p \to \infty$ limit, we can reproduce the computations of Section III, obtaining

$$\lim_{p\to\infty}\mu_p(0) = -\frac{\sup_{s\in[0,N]}\mathsf{B}_N(s) + \inf_{s\in[0,N]}\mathsf{B}_N(s)}{2}. \quad (60)$$

If OBC are considered, then $\mu_p(0) = 0$ and we have simply, $\forall p > 1$,

$$M_p(x) \equiv M(x) = \mathsf{B}_N(x), \quad x \in [0, N]; \quad (61)$$

It can be easily seen that

$$\frac{1}{N}\int_0^N \langle M(x) M(x+r)\rangle\,\mathrm{d}\,x = c(N) + N\phi\left(\frac{r}{N}\right) \quad (62)$$

where $c(N) = \frac{N}{6}$.

V. CONCLUSIONS

In the present work we solved the Euclidean bipartite matching problem on the interval $[0,1]$, using the cost functional (2) in the continuum limit, $N \to \infty$, for $p \to \infty$, both with open boundary conditions and with periodic boundary conditions. The solution is based on the exact correspondence between the optimal map and a Brownian bridge process on the same interval. Moreover, we computed the correlation function for the optimal map and we observed that in all considered cases it has the form

$$\overline{\langle m_p(t) m_p(t+\tau)\rangle} = c_p - \phi(\tau), \quad \tau \in [0,1], \tag{63}$$

for some constant c_p depending on the adopted boundary conditions and on the value p for the optimal cost. Note that if we consider the problem of a matching of N random $\mathcal{B}$-points to N lattice $\mathcal{R}$-points on the interval $[0,N]$, for $s,t \in [0,N]$ the correlation function assumes the form

$$G(\tau) \coloneqq \overline{\langle M_p(t) M_p(t+\tau)\rangle} = c_p N - \frac{|\tau|}{2}\left(1 - \frac{|\tau|}{N}\right), \tag{64}$$

where $M_p(t)$ is the signed distance between $t \in [0,N]$ and its corresponding inverse image under the action of the optimal map. It follows that for $N \to \infty$ the correlation function $G(\tau)$ has a divergent part, $G(0) = c_p N$, depending through c_p on the specific details of the problem (e.g., the boundary conditions adopted or the value of p), a universal finite part $-\frac{|\tau|}{2}$ and a (universal) finite size correction $\frac{\tau^2}{2N}$. We obtained also numerical evidences of the validity of eq. (64) for different values of p both with OBC and with PBC: an exact derivation of c_p for OBC $\forall p > 1$ and for c_2 and c_∞ for PBC was presented. This fact suggests that all Euclidean matching problems in one dimension with strictly convex cost functionals belong to the same universality class and that the specific details of the model determine only the value of the constant c_p in the divergent contribution $G(0)$. Similarly, on the interval $[0,N]$ eq. (32) becomes

$$\begin{aligned}\langle \sigma_p(t)\sigma_p(t+\tau)\rangle &= \\ &= \frac{2}{\pi}\arctan\frac{2c_pN - \frac{\tau}{2}\left(1-\frac{\tau}{N}\right)}{\sqrt{\frac{\tau}{2}\left(1-\frac{\tau}{N}\right)\left(4c_pN - \tau\left(1-\frac{\tau}{N}\right)\right)}} \\ &= 1 - \frac{1}{\pi}\sqrt{\frac{2\tau}{c_pN}} + o\left(\frac{1}{\sqrt{N}}\right),\end{aligned} \tag{65}$$

in which the universal part is given by the constant 1 and finite size corrections scale as $\frac{1}{\sqrt{N}}$ up to a scaling factor depending on c_p.

VI. ACKNOWLEDGEMENTS

The authors are grateful to Luigi Ambrosio, from Scuola Normale Superiore in Pisa, and Andrea Sportiello, from Université Paris 13, for useful and stimulating discussions.

Appendix A: Joint distribution for the solution of the problem with open boundary conditions

In the present section we derive the joint distribution of the matching map m for the solution of the Euclidean bipartite matching problem on the line. In the hypothesis that $0 < t_1 < t_2 < \cdots < t_K < 1$, $K < N$, let us consider K $\mathcal{B}$-points $\{y_{i_k}\}_{k=1,\dots,K}$, $i_1 < i_2 < \cdots < i_K$, and evaluate the following quantity:

$$\begin{aligned}&\Pr\left(\{y_{i_k} \in \mathrm{d}\, t_k\}_{k=1,\dots,K}\right) = \\ &\quad = \frac{N!(1-t_K)^{N-i_K}}{(N-i_K)!}\prod_{k=1}^{K}\frac{(\Delta t_k)^{\Delta i_k - 1}\,\mathrm{d}\, t_k}{(\Delta i_k - 1)!} \\ &\quad = \binom{N}{\Delta i_1 \dots \Delta i_{K+1}}\prod_{k=1}^{K+1}(\Delta t_k)^{\Delta i_k}\prod_{k=1}^{K}\frac{\Delta i_k}{\Delta y_k}\,\mathrm{d}\, t_k\end{aligned} \tag{A1}$$

where $i_0 \coloneqq 0$, $i_{K+1} \coloneqq N$, $t_0 \coloneqq 0$, $t_{K+1} \coloneqq 1$ and

$$\Delta t_k \coloneqq t_k - t_{k-1}, \quad \Delta i_k \coloneqq i_k - i_{k-1}. \tag{A2}$$

Note that the previous equation has the form of a multinomial distribution. Introducing the rescaled variable $m_k(t_k) = \sqrt{N}\left(t_k - \frac{i_k}{N}\right)$, in the large N limit we obtain a multivariate Gaussian distribution in the variable $\{\Delta m_k\}_{k=1,\dots,K}$, $\Delta m_k \coloneqq m_k - m_{k-1}$, $m_0 \coloneqq 0$, whose covariance matrix is given by the degenerate matrix

$$\Sigma_{ij}(\{y\}) = \delta_{ij}\Delta t_i - \Delta t_i \Delta t_j. \tag{A3}$$

Note that $\det\Sigma = \left(1 - \sum_{k=1}^{K+1}\Delta t_k\right)\prod_{k=1}^{K+1}\Delta t_k = 0$: the constraint $\sum_i \Delta t_i = 1$ reduces the rank of the matrix from $K+1$ to K (we have indeed K independent variables). In this case, no density distribution exists in the $(K+1)$-dimensional space of variables $\Delta m_1, \dots, \Delta m_{K+1}$, due to the fact that an additional constraint, $\sum_{k=1}^{K+1}\Delta m_k = 0$, holds, and therefore the distribution is *singular*. We need to restrict therefore to the subspace of K variables $\{\Delta m_k\}_{k=1,\dots K}$. The distribution is still a multivariate Gaussian but with covariance matrix $\Sigma_{ij} = \Sigma_{ij}$ with $i,j = 1,\dots,K$. The covariance matrix is positive definite and non singular on this subspace, being $\det\Sigma = \left(1 - \sum_{k=1}^{K}\Delta t_k\right)\prod_{k=1}^{K}\Delta t_k = \prod_{k=1}^{K+1}\Delta t_k$ and

$$[\Sigma(y)]^{-1}_{ij} = \frac{\delta_{ij}}{\Delta t_i} + \frac{1}{\Delta t_{K+1}} \tag{A4}$$

The joint distribution is therefore

$$\begin{aligned}&\Pr\left(\{m_k(t_k) \in \mathrm{d}\, m_k\}_{k=1,\dots K}\right) = \\ &= \frac{\prod_{k=1}^{K}\mathrm{d}\, m_k}{\sqrt{(2\pi)^K\prod_{k=1}^{K+1}\Delta t_k}}\exp\left(-\sum_{k=1}^{K}\frac{(\Delta m_k)^2}{2\Delta t_k} - \frac{m_K^2}{2\Delta t_{K+1}}\right).\end{aligned} \tag{A5}$$

The previous distribution is exactly the joint distribution for the Brownian bridge process, proving the equivalence of the processes in the large N limit. In particular, the two point joint distribution for the Brownian bridge is given by eq. (A5) for $K = 2$:

$$\Pr\left(m_1(t_1) \in \mathrm{d}\, m_1, m_2(t_2) \in \mathrm{d}\, m_2\right) = \\ = \frac{\mathrm{d}\, m_1\, \mathrm{d}\, m_2}{2\pi\sqrt{(1-t_2)(t_2-t_1)t_1}}\, \mathrm{e}^{-\frac{m_1^2}{2t_1} - \frac{(m_2-m_1)^2}{2(t_2-t_1)} - \frac{m_2^2}{2(1-t_2)}} \tag{A6}$$

where $0 < t_1 < t_2 < 1$ is assumed.

Appendix B: Probability distributions for the Brownian bridge

In the present section we briefly present, without proofs, some fundamental probability results on the Brownian bridge process and on some noteworthy probability distributions related to it. As explained before, the distribution for the sup of the absolute value of a Brownian bridge is given by the well known Kolmogorov distribution, eq. (20). Using the reflection principle and Bayes' theorems, it can be also proved that [13, 16]

$$\Pr\left(\sup_{t\in[0,1]} \mathsf{B}(t) < M, \inf_{t\in[0,1]} \mathsf{B}(t) > -m\right) = \\ = 1 + 2\sum_{k=1}^{\infty} \mathrm{e}^{-2k^2(M+m)^2} + \\ - \sum_{k=0}^{\infty}\left(\mathrm{e}^{-2(km+(k+1)M)^2} + \mathrm{e}^{-2(kM+(k+1)m)^2}\right). \tag{B1}$$

For $t \in (0,1)$ and $M > b$, it can be easily obtained that

$$\Pr\left(\sup_{s\in[0,1]} \mathsf{B}(s) < M, \mathsf{B}(t) \in \mathrm{d}\, b\right) = \\ = \frac{\mathrm{e}^{-\frac{b^2}{2t(1-t)}}\left(1-\mathrm{e}^{-\frac{2M(M-b)}{t}}\right)\left(1-\mathrm{e}^{-\frac{2M(M-b)}{1-t}}\right)}{\sqrt{2\pi t(1-t)}}\, \mathrm{d}\, b. \tag{B2}$$

The computation of (36) can be easily performed using the distribution of the sup of the Brownian bridge

$$\Pr\left(\sup_{t\in[0,1]} \mathsf{B}(t) < M\right) = 1 - \mathrm{e}^{-2M^2}. \tag{B3}$$

[1] H. W. Kuhn, Naval research logistics quarterly **2**, 83 (1955).
[2] J. Munkres, Journal of the Society for Industrial and Applied Mathematics **5**, pp. 32 (1957), ISSN 03684245.
[3] J. Edmonds and R. M. Karp, Journal of the ACM **19**, 248 (1972).
[4] D. J. Aldous, Random Structures and Algorithms pp. 381–418 (2001).
[5] M. Mézard and G. Parisi, Journal de Physique Lettres **46**, 771 (1985).
[6] M. Mézard and G. Parisi, Journal de Physique **49**, 2019 (1988).
[7] E. Boniolo, S. Caracciolo, and A. Sportiello, arXiv (2014), arXiv:1403.1836v1.
[8] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, Phys. Rev. E **90**, 012118 (2014).
[9] T. Leighton and P. W. Shor, Combinatorica **9**, 161 (1989).
[10] P. W. Shor and J. E. Yukich, The Annals of Probability **19**, 1338 (1991).
[11] S. K. Nechaev, A. N. Sobolevski, and O. V. Valba, Physical Review E **87**, 012102 (2013), ISSN 1539-3755.
[12] L. A. Shepp, The Annals of Probability **10**, 234 (1982).
[13] R. Dudley, *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, 2002).
[14] C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften (Springer, 2008), ISBN 9783540710509.
[15] M. Barczy and E. Iglói, Central European Journal of Mathematics **9**, 65 (2010).
[16] L. Beghin and E. Orsingher, Lithuanian Mathematical Journal **39**, 157 (1999).