



HAL
open science

Universality in the one-dimensional euclidean random assignment problem

Yuqi Liu

► **To cite this version:**

Yuqi Liu. Universality in the one-dimensional euclidean random assignment problem. Mathematics [math]. 2021. dumas-03706851

HAL Id: dumas-03706851

<https://dumas.ccsd.cnrs.fr/dumas-03706851v1>

Submitted on 28 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Universality in the one-dimensional Euclidean random assignment problem

Étudiante: Yuqi LIU Encadrant: Matteo D'ACHILLE

Master 1 Mathématiques et applications UE: Travail d'Étude et de Recherche

17/05/2021

Résumé

Cette thèse est composée de quatre chapitres thématiques. Le premier chapitre commence par rappeler quelques notions élémentaires de la théorie des graphes et introduit le problème général d'assignation aléatoire euclidienne. Le deuxième chapitre est une revue de littérature, centrée sur la dimension 1. Le troisième chapitre est consacré à l'étude d'un nouveau problème d'assignation aléatoire euclidienne "unidimensionnel". Le dernier chapitre 4 présente quelques perspectives pour les recherches futures.

Abstract

This thesis is composed of four themed chapters. The first chapter begins by reviewing some elementary notion from the theory of graphs, and introduces the general Euclidean Random Assignment Problem (ERAP). Chapter 2 is a review of the literature, focusing in one dimension. The third chapter is concerned with the study of a new "1-dimensional" Euclidean random assignment problem. The ending chapter 4 presents some perspectives for future research.

1 Introduction

The theory of graphs came into existence during the first half of the 18th century with Euler publishing the paper on the Seven Bridges of Königsberg [1]. This theory is particularly useful in dealing with combinatorial optimization problems, and in particular in the study of linear assignment problems. Thus, before explaining our problem, it is necessary to recall some definitions about graphs. The reference for Subsection 1.1 is Chartrand *et al.* [2].

1.1 Fundamentals of simple graph

A *simple graph* is a graph, without loops or parallel edges, in which every edge is undirected. A simple graph G consists of a finite nonempty set $V(G)$ of objects called *vertices* and a set $E(G)$ of 2-element subsets of $V(G)$ called *edges*. If v_1v_2 is an edge of G , then vertices v_1 and v_2 are said to be *adjacent* in G . The number of vertices and edges in G are called the *order* and *size* of G , and denoted by $n(G)$ and $m(G)$, respectively. The *degree* of a vertex v in a graph G is the number of edges incident with v and is denoted by $\deg_G v$.

Theorem. (The handshake lemma) If G is a simple graph of size m , then

$$\sum_{v \in V(G)} \deg_G v = 2m.$$

A graph G is a *bipartite graph* if $V(G)$ can be partitioned into two subsets U and W , called *partite sets*, such that every edge of G joins a vertex of U and a vertex of W . We call G a *complete bipartite graph* if every vertex of U is adjacent to every vertex of W . A complete bipartite graph with $n(U) = s$ and $n(W) = t$ is denoted by $K_{s,t}$ or $K_{t,s}$, so that its order is $s + t$ and its size is $s \times t$.

A collection of adjacent vertices that repeats no vertex (circuit), except for the first and last, is a *cycle*. A k -*cycle* is a cycle of length k . Graph $K_{5,3}$ and 6-*cycle* are shown in Figure 1.1 and Figure 1.2.

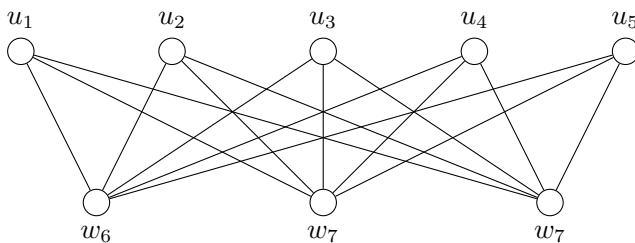


Figure 1.1: Complete bipartite graph $K_{5,3}$

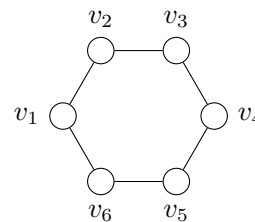


Figure 1.2: 6-cycle C_6

Various matrices can be used to study graphs. Let G be a simple graph of order n and size m , where $V(G) = \{v_1, \dots, v_n\}$ and $E(G) = \{e_1, e_2, \dots, e_m\}$. The *adjacency matrix* of G is the $n \times n$ matrix

$A = (a_{ij})$, where

$$a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in E(G) \\ 0 & \text{otherwise,} \end{cases}$$

while the *incidence matrix* of G is the $n \times m$ matrix $B = (b_{ij})$, where

$$b_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is incident with } e_j \\ 0 & \text{otherwise.} \end{cases}$$

We say that $G' = (V, E')$ is a *spanning subgraph* or a *factor* of $G = (V, E)$ if $E' \subset E$. If every vertex of a factor has degree r then we call it an *r-factor*. A *matching* in $K_{n,n}$ is an 1-factor of $K_{n,n}$.

If a simple graph G has no cycles, it is called *acyclic* or a *forest*. A *tree* is an acyclic connected simple graph. The size of a tree of order n is $n - 1$.

1.2 The linear assignment problem

Let us now consider a combinatorial optimization problem, which is called “the assignment problem”, and formulate it in the language developed in the previous section. (Appendix A provides a brief overview of combinatorial optimization.) In a complete bipartite graph $K_{n,n}$ with partite sets $B = \{b_1, b_2, \dots, b_n\}$ (blue dots in Figure 1.3) and $R = \{r_1, r_2, \dots, r_n\}$ (red dots in Figure 1.3), given a positive valued¹ cost function f defined on the edges, $f : E \rightarrow \mathbb{R}^+$, we want to find a matching $T = (V, E')$ of G for which the objective function

$$f(T) = \sum_{1 \leq i, j \leq n} a_{ij} f(b_i r_j) = \sum_{b_i r_{\pi(i)} \in E'} f(b_i r_{\pi(i)})$$

is minimal. In this linear programming problem (see Appendix A), $(a_{ij})_{n \times n}$ is the adjacency matrix of T . The set of possible matchings is in bi-jection with the symmetric group \mathcal{S}_n , i.e. the group of all permutations of a set with n elements. The number of possible matchings is $n!$.

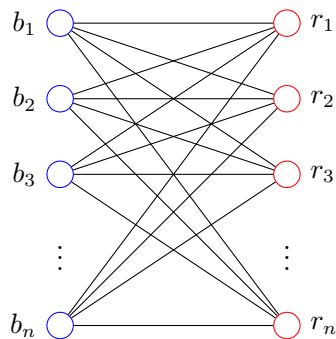


Figure 1.3: Graph $K_{n,n}$

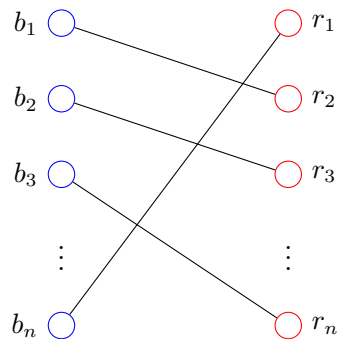


Figure 1.4: A matching of $K_{n,n}$

¹There is no loss of generality in this assumption, as a lower-bounded cost function f is enough for our scopes.

1.3 The Euclidean Random Assignment Problem

Turning now to formulate the d -dimensional Euclidean Random Assignment Problem (or “ERAP” for brevity) following the discussion in [3]. This problem was first articulated in statistical mechanics by Mézard and Parisi as a prototypical model of finite-dimensional spin glass [4].

In statistical mechanics, a primary concern are the typical properties of a system composed of many particles in the limit of low temperature. As the temperature is lowered the Boltzmann distribution (also called Gibbs distribution) collapses into the lowest energy state, which would dominate the properties of matter [5].

Let us expand a bit on this remark with the following observations. Consider a statistical mechanical system with energies $\{E_1, \dots, E_n\}$ in bijection with the microscopic configurations. The canonical partition function of the system [6, 7] at inverse temperature β is

$$Z_n(\beta) = \sum_{i=1}^n e^{-\beta E_i},$$

and the associated free energy is

$$f_n(\beta) = -\frac{1}{\beta} \log Z_n(\beta).$$

Since

$$e^{\max(-\beta E_1, -\beta E_2, \dots, -\beta E_n)} \leq \sum_{i=1}^n e^{-\beta E_i} \leq n e^{\max(-\beta E_1, -\beta E_2, \dots, -\beta E_n)},$$

on taking logs we obtain

$$\begin{aligned} \max(-\beta E_1, -\beta E_2, \dots, -\beta E_n) &\leq \log \sum_{i=1}^n e^{-\beta E_i} \leq \max(-\beta E_1, -\beta E_2, \dots, -\beta E_n) + \log n, \\ -\frac{1}{\beta} \max(-\beta E_1, -\beta E_2, \dots, -\beta E_n) &\geq -\frac{1}{\beta} \log \sum_{i=1}^n e^{-\beta E_i} \geq -\frac{1}{\beta} \max(-\beta E_1, -\beta E_2, \dots, -\beta E_n) - \frac{1}{\beta} \log n. \end{aligned}$$

Indeed,

$$-\frac{1}{\beta} \max(-\beta E_1, -\beta E_2, \dots, -\beta E_n) = \min(E_1, E_2, \dots, E_n),$$

$$\lim_{\beta \rightarrow +\infty} \frac{1}{\beta} \log n = 0.$$

Then, we have

$$\lim_{\beta \rightarrow +\infty} f_n(\beta) = \lim_{\beta \rightarrow +\infty} -\frac{1}{\beta} \log \sum_{i=1}^n e^{-\beta E_i} = \min(E_1, E_2, \dots, E_n). \quad (1)$$

Therefore, one could hope to use the properties of the ground state energy at zero temperature (left side of equation (1)) depending on some choices of disorder to understand the properties of the solution of the stochastic version of a combinatorial optimization problem (in which E_1, E_2, \dots, E_n represent the possible values taken by the objective function of interest).

The ferromagnetic bonds and the antiferromagnetic bonds in a spin glass are sampled from an instance of a disorder distribution related to distances among “atoms” or “impurities” [8], which are represented here by B and R . In particular, they are represented by points in a d -dimensional metric space (Ω, D) . The statistical ensemble for the disorder distribution is denoted by (v_B, v_R) . Playing the role of energy, the cost function is restricted to a power function with an exponent p :

$$f(b_i r_{\pi(i)}) = D^p(b_i, r_{\pi(i)}),$$

in order to share basic requirements for a physical system at criticality. For $2n = n + n$ points, we get a $n \times n$ assignment cost matrix

$$c = \begin{pmatrix} D^p(b_1, r_1) & D^p(b_1, r_2) & \dots & D^p(b_1, r_n) \\ D^p(b_2, r_1) & D^p(b_2, r_2) & \dots & D^p(b_2, r_n) \\ \vdots & \vdots & \ddots & \vdots \\ D^p(b_n, r_1) & D^p(b_n, r_2) & \dots & D^p(b_n, r_n) \end{pmatrix},$$

and $n!$ permutation matrices P_π , with $P_{i,j} = \delta_{j,\pi(i)}$.

Here, $D(b_i, r_j)$ is almost exclusively d -dimensional Euclidean distance of two vertices. But other choices for the metric D are possible.

A specific ERAP, built into this system, is completely specified by the triple $((\Omega, D), (v_B, v_R), p)$. The energy (Hamiltonian) of a configuration π is

$$\mathcal{H}^{(p)}(\pi) = \sum_{i=1}^n D^p(b_i, r_{\pi(i)}) = \text{Tr}[P_\pi c].$$

Then we will consider the optimal assignment π_{opt} and the optimal cost

$$\mathcal{H}_{\text{opt}}^{(p)} := \mathcal{H}^{(p)}(\pi_{\text{opt}}) = \min_{\pi \in \mathcal{S}_n} \mathcal{H}^{(p)}(\pi).$$

The remaining part of this memoir will focus mostly on one-dimensional Euclidean random assignment problem, which means that B and R are distributed in a 1-dimensional metric space.

1.4 The link with Monge–Kantorovich problem

The relationship between two problems that both play an important role in optimal transport theory: the Euclidean random assignment problem and the Monge–Kantorovich problem, has already been explored. Brezis [9] has recently remarked that what we call \mathcal{H}_{opt} in the discrete setting of ERAP is proportional to the so-called Wasserstein distance among the empirical distributions associated to Blue and Red points in the continuous setting of Monge–Kantorovich problem. Thus, when we discuss universality in the one dimensional ERAP, we can refer to the conclusions in Monge–Kantorovich

problem, and some remarkable connections emerge. For example, the transport field in the one dimensional ERAP on the unit interval for $p \geq 1$ has the Brownian bridge as a scaling limit [10].

Vershik *et al.* [11] have reminded of the history of the Wasserstein distance and the Monge–Kantorovich problem, and the remarkable role of them in linear programming. Recall that the Wasserstein transport distance $W_p(\mu_n, \mu)$, $p \geq 1$, between μ_n and μ is defined by

$$W_p^p(\mu_n, \mu) = \inf_{\pi} \int_{\mathbf{R}} \int_{\mathbf{R}} d(x-y)^p d\pi(x, y) = \mathbb{E} \left[\frac{\mathcal{H}_{\text{opt}}^{(p)}}{n} \right].$$

This definition in [12] indicates the link between $W_p(\mu_n, \mu)$ and $\mathcal{H}_{\text{opt}}^{(p)}$.

2 Literature review

This chapter is subdivided into two sections according to the value of the energy-distance exponent p . We will recall some results for the case $p > 1$, which will be useful in the next chapter. Then we will briefly mention the previous studies for the case $0 < p < 1$ (concave regime) and the latest result for the borderline case $p = 1$.

2.1 Case $p > 1$

2.1.1 On the configuration π_{opt}

Let the two sets of points be labeled in increasing order according to their position along the line. The following property of the optimal matching has been reported in previous studies [13, 14].

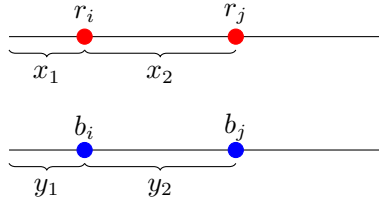
Proposition 1. For $p > 1$,

$$\mathcal{H}_{\text{opt}}^{(p)} = \mathcal{H}^{(p)}(\pi),$$

if and only if

$$\pi(i) = i \text{ for all } 1 \leq i \leq n.$$

Proof. Let the positions of any two red dots r_i, r_j and any two blue dots b_i, b_j be $x_1, x_1 + x_2, y_1, y_1 + y_2$, respectively. If $1 \leq i < j \leq n$, we have $x_2 \geq 0$ and $y_2 \geq 0$.



Depending on the permutations of b_i and b_j , there are two different costs:

$$\left(\begin{array}{cccccccc} 1 & 2 & \dots & i & \dots & j & \dots & n-1 & n \\ \pi(1) & \pi(2) & \dots & i & \dots & j & \dots & \pi(n-1) & \pi(n) \end{array} \right) : \mathcal{H}_1^{(p)} = |x_1 - y_1|^p + |x_1 + x_2 - y_1 - y_2|^p + C,$$

$$\left(\begin{array}{cccccccc} 1 & 2 & \dots & i & \dots & j & \dots & n-1 & n \\ \pi(1) & \pi(2) & \dots & j & \dots & i & \dots & \pi(n-1) & \pi(n) \end{array} \right) : \mathcal{H}_2^{(p)} = |y_1 + y_2 - x_1|^p + |x_1 + x_2 - y_1|^p + C.$$

We need to prove that

$$|d - y_2|^p + |d + x_2|^p \geq |d|^p + |d + x_2 - y_2|^p, \text{ with } d = x_1 - y_1.$$

For $p > 1$, convexity of $f(x) = |x|^p$ gives

$$f(d + x_2) - f(d) > f(d + x_2 - y_2) - f(d - y_2).$$

This completes the proof. □

2.1.2 ERAP and Brownian bridge

Let $\{b_i, 1 \leq i \leq n\}$ and $\{r_i, 1 \leq i \leq n\}$ be $2n$ uniform i.i.d. random variables on the interval $[0, 1]$. For $n \rightarrow +\infty$, a link between the ERAP and the Brownian bridge has been shown using Donsker's theorem [15].

Theorem. (Donsker's) Let $\{x_i, i = 1, \dots, n\}$ be n independent observations sampled from a distribution function $F(x)$, and let $F_n(x)$ be the empirical cumulative function

$$F_n(x) := \frac{1}{n} \{\#x_i | x_i < x\}.$$

Then, as $n \rightarrow +\infty$, $\sqrt{n}(F_n(x) - F(x))$ converges in distribution to the Brownian Bridge B_x^{br} .

The standard Brownian Bridge B_x^{br} is a centered Gaussian process with covariance $\mathbb{E}(B_t^{br} B_s^{br}) = t(1-s)$ when $0 \leq t \leq s \leq 1$ (see [10] for a review and [16] for remarks on the relevance of the Brownian Bridge in one dimensional stochastic models).

The idea of a *Grid-Poisson ERAP* was first proposed in Boniolo *et al.* (2014) [13] and Caracciolo *et al.* (2014) [17]. In this problem, B , replaced by P_{grid} , is a deterministic grid on the domain Λ_n rather than a set of independent random variables. A common choice has been the uniform grid:

$$P_{grid} = \{p_i^{grid} | p_i^{grid} = \frac{i}{n+1}, 1 \leq i \leq n\}.$$

From Donsker's theorem, we have (see Figure B.1(1) for a pictorial representation)

$$\sqrt{n}(p_i^{grid} - r_i) \xrightarrow{n \rightarrow +\infty} B^{br}(s; \omega).$$

In the same way, (shown in Figure B.1(2))

$$\sqrt{n}(b_i - p_i^{grid}) \xrightarrow{n \rightarrow +\infty} -B^{br}(s; \tilde{\omega}).$$

Denoting by

$$\varphi_i := b_i - r_i, \text{ for all } 1 \leq i \leq n,$$

the authors of Refs. [14] have presented that (see Figure B.1(3))

$$\phi(s) := \sqrt{n}\varphi_{ns+\frac{1}{2}} \xrightarrow{n \rightarrow +\infty} B^{br}(s; \omega) - B^{br}(s; \tilde{\omega}),$$

here, s is a new variable such that $i = ns + \frac{1}{2}$, $B^{br}(s; \omega)$ and $B^{br}(s; \tilde{\omega})$ are independent.

This has allowed the computation of some statistical properties of the minimal energy. For example, at $p = 2$, we have

$$\begin{aligned}
& \lim_{n \rightarrow +\infty} \mathbb{E} [\mathcal{H}_{\text{opt}}^{(2)}] \\
&= \lim_{n \rightarrow +\infty} \mathbb{E} [\mathcal{H}^{(2)}(\pi_{\text{opt}})] \\
&= \lim_{n \rightarrow +\infty} \mathbb{E} \left[\sum_{i=1}^n |\varphi_i|^2 \right] \\
&= \mathbb{E} \left[\lim_{n \rightarrow +\infty} \sum_{i=1}^n |\varphi_i|^2 \right] \\
&= \mathbb{E} \left[\int_0^1 \phi(s)^2 ds \right]
\end{aligned}
\quad \left| \quad \begin{aligned}
&= \mathbb{E} \left[\int_0^1 (B^{br}(s; \omega) - B^{br}(s; \tilde{\omega}))^2 ds \right] \\
&= \mathbb{E} \left[\int_0^1 (B^{br}(s; \omega))^2 ds \right] + \mathbb{E} \left[\int_0^1 (B^{br}(s; \tilde{\omega}))^2 ds \right] \\
&= 2 \int_0^1 s(1-s) ds \\
&= 2 \left[\frac{s^2}{2} - \frac{s^3}{3} \right]_0^1 \\
&= \frac{1}{3},
\end{aligned}$$

(a similar computation holds for any $p > 1$). More recently, due to the an exact rewriting of $\mathbb{E}[\mathcal{H}_{\text{opt}}^{(p)}]$ in terms of a Generalized Selberg Integrals [18], it has been shown that there exists an exact formula valid

$$\forall n, \text{ and } \forall p \geq 1, \mathbb{E}[\mathcal{H}_{\text{opt}}^{(p)}] = n \frac{\Gamma(1 + p/2)\Gamma(n + 1)}{(p + 1)\Gamma(n + 1 + p/2)},$$

where Γ is Euler's function.

2.2 Case $0 < p \leq 1$

Whilst this case has received much less attention than the case $p > 1$, it did have several practical applications in economic terms. Imagine being in a railway line or highway, the concavity of the cost function can reflect a shipping tariff that increases with the distance, even while the cost per mile shipped goes down. McCann [19] has introduced the application model and the following proposition.

Proposition 2. (No-crossing rule) For $0 < p < 1$, π_{opt} has no crossing.



Figure 2.1: No-crossing rule

There is usually more than one assignment that satisfies the non-crossing property, so this property is not sufficient to fully characterize the optimal assignment. Recently, a specific matching rule, giving a canonical but sub-optimal configuration, the Dyck matching, has been proposed [20]. On the basis of extensive numerical evidences, it has been conjectured that the expected energy of Dyck matchings has the same leading asymptotics of the unknown optimal matching, up to constants in n [20].

The case $p = 1$ has also attracted some interest recently [21], as it is degenerate. As the cost function changes concavity at $p = 1$, there are at least two distinct optimal matchings for each instance, the ordered matching and the Dyck matching. It is atypical for these two situations to coincide.

For the transportation problem, the case of the line segment is idealized, and in the next chapter we investigate how these conclusions can be extended to a crossroads.

3 A 1-dimensional Euclidean random assignment problem

3.1 Hypothesis

We begin with presenting the triple $((\Omega, D), (v_B, v_R), p)$ in this “one-dimensional” ERAP.

A labeled tree G^* with k edges is illustrated in Figure 3.1, k is a positive integer. n blue dots $\{b_i, 1 \leq i \leq n\}$ and n red dots $\{r_i, 1 \leq i \leq n\}$ are distributed in these k edges according to the following law:

- For a dot v , the distance from center is denoted by $v^{(1)}$, and follows a continuous uniform variable on the interval $[0, \frac{1}{k}]$: $v^{(1)} \sim \mathcal{U}(0, \frac{1}{k})$;
- For a dot v , the direction (i.e. the edge which a dot is on) is denoted by $v^{(2)}$, and follows a discrete uniform distribution: $v^{(2)} \in \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ and $\mathbb{P}(v^{(2)} = \alpha_i) = \frac{1}{k}$, for all $1 \leq i \leq k$;
- The dot v is labeled as $(v^{(1)}, v^{(2)})$.

For $k = 3$, a realization of 2 blue dots $\{b_1, b_2\}$ and 2 red rods $\{r_1, r_2\}$ is illustrated in Figure 3.2.

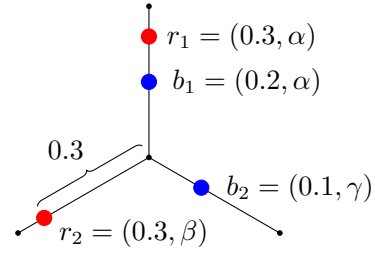
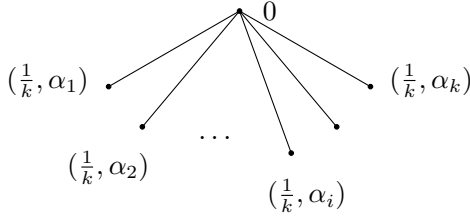


Figure 3.1: A labeled tree G^* with k edges

Figure 3.2: An example of distribution on G^*

For two dots $v_1 = (v_1^{(1)}, v_1^{(2)})$ and $v_2 = (v_2^{(1)}, v_2^{(2)})$, the distance (see Figure 3.3) between them is defined as following:

$$D(v_1, v_2) = \begin{cases} |v_1^{(1)} - v_2^{(1)}|, & \text{if } v_1^{(2)} = v_2^{(2)}, \text{ i.e. they are in the same direction,} \\ v_1^{(1)} + v_2^{(1)}, & \text{if } v_1^{(2)} \neq v_2^{(2)}, \text{ i.e. they are not in the same direction.} \end{cases}$$

We will consider the case $p \geq 1$ in the section 3.2, and the case $p = 2$ in the section 3.3.

3.2 On the configuration π_{prune}

Inspired by the work of Boniolo *et al.* [13] and of Caracciolo *et al.* [20], we will introduce a sub-optimal configuration, which is called π_{prune} . On this configuration, we are able to build a unique solution of $\mathcal{H}^{(p)}$, then show in the next section that this solution is close to the optimal solution. The following steps will be exemplified for the case $k = 3$ but generalize simply to other values of k .

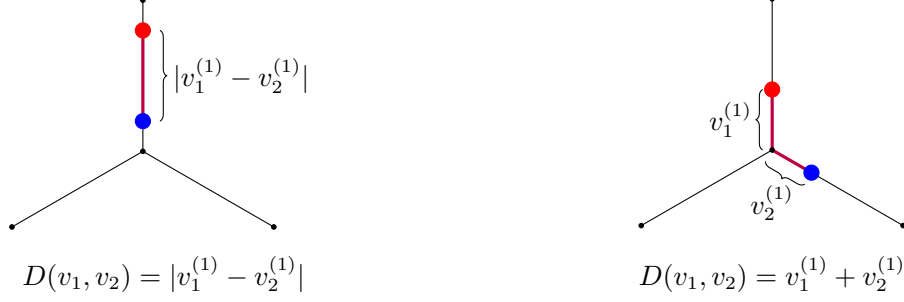


Figure 3.3: Distance between two dots

Step 1.

Suppose there are m_b^α blue dots and m_r^α red dots with direction² α , they are :

- $\{b_i = (b_i^{(1)}, \alpha) : b_i^{(1)} \geq b_j^{(1)}, 1 \leq i \leq j \leq m_b^\alpha \leq n\}$ and
- $\{r_i = (r_i^{(1)}, \alpha) : r_i^{(1)} \geq r_j^{(1)}, 1 \leq i \leq j \leq m_r^\alpha \leq n\}$.

For the $m^\alpha (= \min(m_r^\alpha, m_b^\alpha))$ pairs of data, we define $\mathcal{H}_\alpha^{(p)}(\pi)$, which is a part of $\mathcal{H}^{(p)}(\pi)$, as follows:

$$\mathcal{H}_\alpha^{(p)}(\pi) = \sum_{i=1}^{m^\alpha} D^p(b_i, r_{\pi(i)}), \text{ for all } p \geq 1,$$

and the corresponding optimal cost

$$\mathcal{H}_{\alpha, \text{opt}}^{(p)} = \min_{\pi \in \mathcal{S}_{m^\alpha}} \mathcal{H}_\alpha^{(p)}(\pi), \text{ for all } p \geq 1.$$

According to **Proposition 1**, for $p \geq 1$, if we neglect the remaining points,

$$\mathcal{H}_{\alpha, \text{opt}}^{(p)} = \mathcal{H}_\alpha^{(p)}(\pi),$$

if and only if

$$\pi(i) = i \text{ for all } 1 \leq i \leq m^\alpha.$$

For the convenience of follow-up discussion, we delete the labels of some dots: $\{b_i, r_i : i > m^\alpha\}$.

Step 2.

In the same way, we suppose there are m_b^β blue dots and m_r^β red dots with direction β , they are :

- $\{b_i = (b_i^{(1)}, \beta) : b_i^{(1)} \geq b_j^{(1)}, m^\alpha < i \leq j \leq m^\alpha + m_b^\beta \leq n\}$ and
- $\{r_i = (r_i^{(1)}, \beta) : r_i^{(1)} \geq r_j^{(1)}, m^\alpha < i \leq j \leq m^\alpha + m_r^\beta \leq n\}$.

²For simplicity, we replace $\{\alpha_1, \alpha_2, \alpha_3\}$ with $\{\alpha, \beta, \gamma\}$.

After defining $m^\beta = \min(m_r^\beta, m_b^\beta)$, we delete the labels of some dots: $\{b_i, r_i : i > m^\alpha + m^\beta\}$. Similarly, we label m_b^γ blue dots and m_r^γ red dots with direction γ :

- $\{b_i = (b_i^{(1)}, \gamma) : b_i^{(1)} \geq b_j^{(1)}, m^\alpha + m^\beta < i \leq j \leq m^\alpha + m^\beta + m_b^\gamma \leq n\}$ and
- $\{r_i = (r_i^{(1)}, \gamma) : r_i^{(1)} \geq r_j^{(1)}, m^\alpha + m^\beta < i \leq j \leq m^\alpha + m^\beta + m_r^\gamma \leq n\}$.

Then, we define $m^\gamma = \min(m_r^\gamma, m_b^\gamma)$, and delete the labels of some dots: $\{b_i, r_i : i > m^\alpha + m^\beta + m^\gamma\}$. For all $p \geq 1$, $\mathcal{H}_\beta^{(p)}(\pi)$, $\mathcal{H}_\gamma^{(p)}(\pi)$, $\mathcal{H}_{\beta, \text{opt}}^{(p)}$ and $\mathcal{H}_{\gamma, \text{opt}}^{(p)}$ are defined as follows:

$$\mathcal{H}_\beta^{(p)}(\pi) = \sum_{i=m^\alpha+1}^{m^\alpha+m^\beta} D^p(b_i, r_{\pi(i)}), \quad \mathcal{H}_\alpha^{(p)}(\pi) = \sum_{i=m^\alpha+m^\beta+1}^{m^\alpha+m^\beta+m^\gamma} D^p(b_i, r_{\pi(i)}),$$

$$\mathcal{H}_{\beta, \text{opt}}^{(p)} = \min_{\pi \in \mathcal{S}_{m^\beta}} \mathcal{H}_\beta^{(p)}(\pi), \quad \mathcal{H}_{\gamma, \text{opt}}^{(p)} = \min_{\pi \in \mathcal{S}_{m^\gamma}} \mathcal{H}_\gamma^{(p)}(\pi).$$

Likewise, according to **Proposition 1**, for $p \geq 1$,

$$\mathcal{H}_{\beta, \text{opt}}^{(p)} = \mathcal{H}_\beta^{(p)}(\pi), \quad \mathcal{H}_{\gamma, \text{opt}}^{(p)} = \mathcal{H}_\gamma^{(p)}(\pi),$$

if and only if

$$\pi(i) = i \text{ for all } m^\alpha + 1 \leq i \leq m^\alpha + m^\beta + m^\gamma.$$

An example of optimal solution in each direction is illustrated in Figure 3.4.

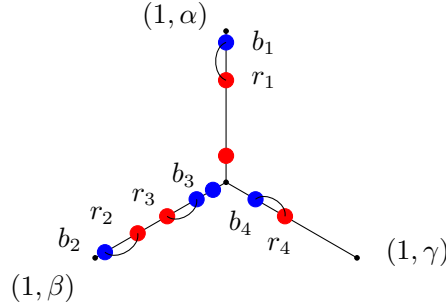


Figure 3.4: Connection in each direction

In the first two steps, if we swap the order of α , β and γ , the result remains the same.

Step 3.

After the second step, there are m^* blue dots and m^* red dots left unlabelled. Here, $m^* = n - (m^\alpha + m^\beta + m^\gamma)$.

We label these $2m^*$ dots as follows:

- $\{b_i = (b_i^{(1)}, b_i^{(2)}) : b_i^{(1)} \geq b_j^{(1)}, m^\alpha + m^\beta + m^\gamma < i \leq j \leq n\}$ and

- $\{r_i = (r_i^{(1)}, r_i^{(2)}) : r_i^{(1)} \leq r_j^{(1)}, m^\alpha + m^\beta + m^\gamma < i \leq j \leq n\}$.

For all $p \geq 1$, $\mathcal{H}^{*(p)}(\pi)$ and $\mathcal{H}_{\text{opt}}^{*(p)}$ are defined as follows:

$$\mathcal{H}^{*(p)}(\pi) = \sum_{i=m^\alpha+m^\beta+m^\gamma+1}^n D^p(b_i, r_{\pi(i)}), \quad \mathcal{H}_{\text{opt}}^{*(p)} = \min_{\pi \in \mathcal{S}_{m^*}} \mathcal{H}^{*(p)}(\pi).$$

Figure 3.5 illustrates that, according to **Proposition 1**, for $p \geq 1$,

$$\mathcal{H}_{\text{opt}}^{*(p)} = \mathcal{H}^{*(p)}(\pi),$$

if and only if

$$\pi(i) = i \text{ for all } m^\alpha + m^\beta + m^\gamma + 1 \leq i \leq n.$$

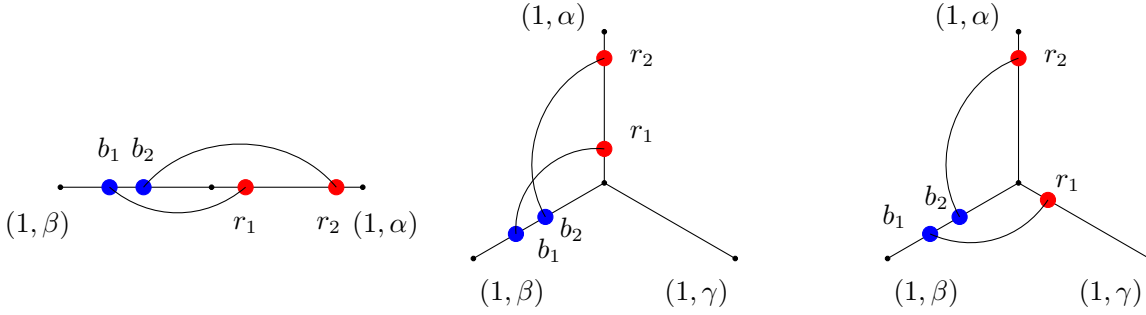


Figure 3.5: The optimal cost is the same in these three cases

In summary, we have defined a matching rule in a canonical way which proceeds through sorting of blue dots and red dots on the branches (“*pruning of the tree*”), and then deals with the remaining points in a unique way. Clearly, this rule generically returns a sub-optimal configuration. The cost of this rule is denoted by $\mathcal{H}_{\text{prune}}^{(p)}$, i.e. $\mathcal{H}_{\text{prune}}^{(p)} = \mathcal{H}_{\alpha, \text{opt}}^{(p)} + \mathcal{H}_{\beta, \text{opt}}^{(p)} + \mathcal{H}_{\gamma, \text{opt}}^{(p)} + \mathcal{H}_{\text{opt}}^{*(p)}$.

3.3 Conclusions

The following conclusions can be drawn from our numerical results:

1. As shown in Figure B.2, we can plot the ratio of the optimal solution³ $\mathcal{H}_{\text{opt}}^{(2)}$ and $\mathcal{H}_{\text{prune}}^{(2)}$ as a function of $n^{-0.5}$. And from the Figure B.3 we can see that the solution $\mathcal{H}_{\text{prune}}^{(2)}$ appears to be close to the optimal solution $\mathcal{H}_{\text{opt}}^{(2)}$.
2. The sum of costs of connection in different directions (i.e. $\mathcal{H}_{\text{opt}}^{*(2)}$) appears to tend to 0, when $n \rightarrow +\infty$ (see Figure B.4).

³We obtained the optimal result in Python with the help of the Jonker-Volgenant algorithm [22], a primal-dual algorithm based on the celebrated *Hungarian algorithm* by Kuhn [23]. The time complexity of the Jonker-Volgenant algorithm is $\mathcal{O}(n^3)$.

3. The ratio of the number of connections in each direction to all (i.e. $\frac{m^{\alpha_i}}{n}$) tend to $\frac{1}{k}$, when $n \rightarrow +\infty$ (see Figure B.5).
4. For the connections in each direction, the distances are related to the Brownian motion (see Figure B.6), that is similar to the conclusion in the section **2.1.2**.

In summary, we can formulate the following conjecture:

Conjecture 1. The expected ground state energy at $p = 2$ is asymptotically

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\mathcal{H}_{\text{opt}}^{(2)} \right] = \frac{3k - 2}{3k^2}.$$

The validity of equation (2) is conformed by the numerical results reported in Figure B.7. Assuming the preceding conclusions, we can get the limits by computing the variance of the standard Brownian bridge stopped at time $1/k$ as follows:

$$\begin{aligned}
\lim_{n \rightarrow +\infty} \mathbb{E} \left[\mathcal{H}_{\text{opt}}^{(2)} \right] &= \lim_{n \rightarrow +\infty} \mathbb{E} \left[\mathcal{H}_{\text{prune}}^{(2)} \right] \\
&= \lim_{n \rightarrow +\infty} \mathbb{E} \left(\sum_{i=1}^{n-m^*} |\varphi_i|^2 \right) \\
&= k \lim_{\alpha_j \rightarrow +\infty} \mathbb{E} \left[\sum_{i=1}^{m^{\alpha_j}} |\varphi_i|^2 \right] \\
&= k \mathbb{E} \left[\lim_{\alpha_j \rightarrow +\infty} \sum_{i=1}^{m^{\alpha_j}} |\varphi_i|^2 \right] \\
&= k \mathbb{E} \left[\int_0^{\frac{1}{k}} \phi(s)^2 ds \right]
\end{aligned}
\quad \left| \quad \begin{aligned}
&= k \mathbb{E} \left[\int_0^{\frac{1}{k}} (B^{br}(s; \omega) - B^{br}(s; \tilde{\omega}))^2 ds \right] \\
&= k \mathbb{E} \left[\int_0^{\frac{1}{k}} (B^{br}(s; \omega))^2 ds \right] + k \mathbb{E} \left[\int_0^{\frac{1}{k}} (B^{br}(s; \tilde{\omega}))^2 ds \right] \\
&= 2k \int_0^{\frac{1}{k}} (s(1-s)) ds \\
&= 2k \left[\frac{s^2}{2} - \frac{s^3}{3} \right]_0^{\frac{1}{k}} \\
&= \frac{3k - 2}{3k^2}.
\end{aligned}
\right. \tag{2}$$

4 Research perspectives

In the previous chapter we have started to study some of the properties on the case of the line segment to crossroads with k branches. Several questions still remain to be answered.

Firstly, for equation (2), we expect a similar result to hold for general $p > 1$, suggesting the following generalization of Conjecture 1

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\mathcal{H}_{\text{opt}}^{(p)}] = 2k \int_0^{\frac{1}{k}} s^{p/2} (1-s)^{p/2} ds.$$

Secondly, although the configuration we employed in section 3.2 greatly reduces the time complexity (from $\mathcal{O}(n^3)$ to $\mathcal{O}(n \log n)$), we could still do better by looking for the *exact* solution among the points falling in the (asymptotically negligible) ball (Figure 4.1) via e.g. the Hungarian algorithm. This would give a new configuration in $\mathcal{O}(\sqrt{n^3})$ complexity, the property of which (such as possible excluded patterns) could be worth studying.

Regarding $\mathcal{H}_{\text{opt}}^{*(2)}$ in section 3.2, the following two properties suggested by our experiments await systematization by rigorous proofs:

- The number of connections in different directions, i.e. m^* (see Figure B.8);
- The maximum distance of these $2m^*$ points from the center, i.e. r_{max} (see Figure B.9).

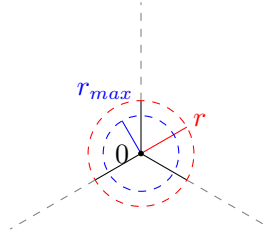


Figure 4.1: Inspiration from π_{prune} to π_{opt}

Finally, the issue of the extension of this problem into higher dimensions is an intriguing one which could be usefully explored in further research. Here we propose two constructions:

- the points of both colors are uniformly distributed on a regular polygon of area 1 with k sides, the circumradii are impassable radii so that matchings among different elementary triangles are realised via the barycenter (see Figure 4.2);
- the points of both colors are uniformly distributed on k semicircles, the diameters of which all coincide (see Figure 4.3).

Due to time constraints, the number of simulated instances in this memoir were only 1000 or 500. Although for one-dimensional problems we started to see a way towards possible reductions in computational complexity (and hence much faster algorithms and statistical power), in higher dimensions a larger number of simulations may be required.

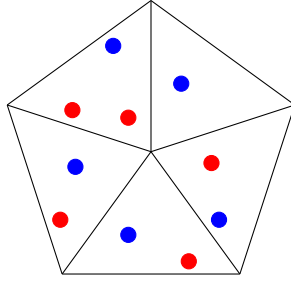


Figure 4.2: Points distributed on a regular pentagon of area 1

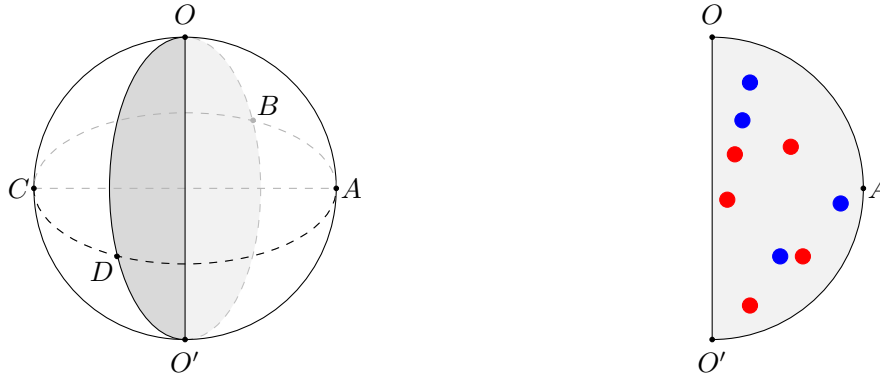


Figure 4.3: Left: 4 semicircles with OO' as diameter. Right: points distributed on a semicircle.

A Combinatorial optimization

This appendix gives a brief review of combinatorial optimization (see [24] for an introductory textbook). Combinatorial optimization concerns problems for which the number of possible solutions is finite and often large. The following is a brief description of the general minimization problem of combinatorial optimization.

Let

$$f : D \rightarrow \mathbf{R}$$

be a real-valued function with domain D . Let

$$S \subseteq D$$

be the set of possible solutions, $x \in S$ is called a feasible solution. The purpose of the general minimization problem of combinatorial optimization is to find $x^* \in S$ such that

$$f(x^*) \leq f(x) \text{ for all } x \in S.$$

The definition of the general maximization problem of combinatorial optimization is analogous.

Typical combinatorial optimization problems are the travelling salesman problem, the transportation problem, the vehicle scheduling problem.

Having introduced what is meant by combinatorial optimization, it is now necessary to move on to discuss some tools for studying these problems.

Since S is finite, we could theoretically solve the general minimization problem with simply exhaustive enumeration as follows:

Evaluate $f(x)$ for all $x \in S$, compare all the function values and choose the smallest.

For many realistic problems the set S have many elements, and it is obvious that the usefulness of enumeration is limited by the size of S . Consequently, different methods have been proposed. For example, the transportation problem can be solved by the stepping-stone method.

The linear programming plays a key role in addressing the issue of combinatorial optimization. We can formulate the linear programming problem in terms of matrices as

$$\text{Max } CX$$

subject to

$$AX \leq b, \text{ and } X \geq 0.$$

Here,

$$C = (c_1, c_2, \dots, c_n), \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad A = (a_{ij})_{m \times n}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

B Supplementary figures

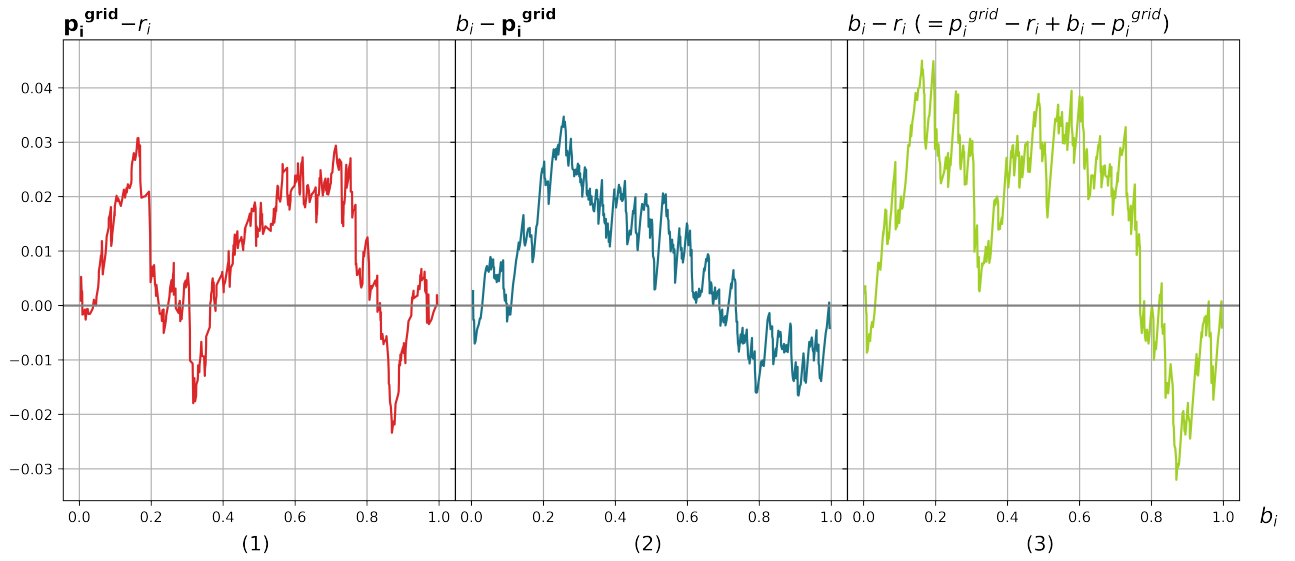


Figure B.1: Optimal transport field for the ERAP with $p = 2$ and $n = 500$

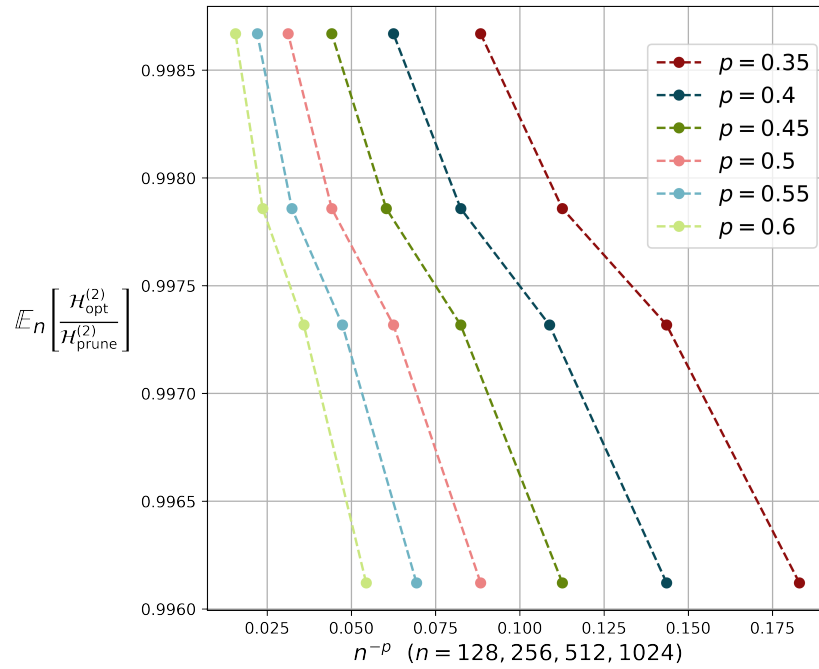


Figure B.2: $\mathbb{E}_n[\mathcal{H}_{\text{opt}}^{(2)}/\mathcal{H}_{\text{prune}}^{(2)}]$ as a function of n^{-p} when $k = 3$. The number of simulated instances is 1000.

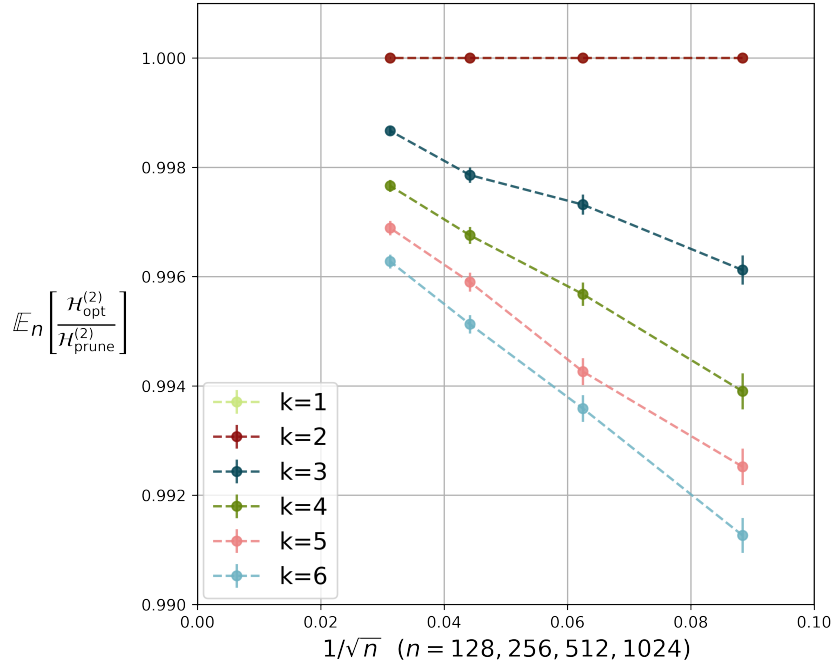


Figure B.3: $\mathbb{E}_n[\mathcal{H}_{\text{opt}}^{(2)}/\mathcal{H}_{\text{prune}}^{(2)}]$ as a function of $1/\sqrt{n}$ with an SEM error bars. The number of simulated instances is 1000. Lines of $k = 1$ and $k = 2$ coincide.

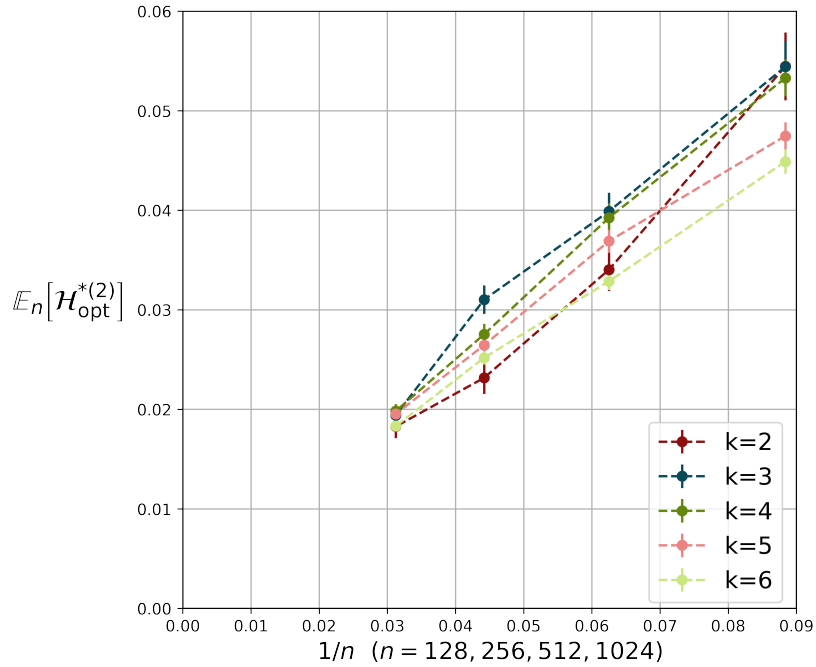


Figure B.4: $\mathbb{E}_n[\mathcal{H}_{\text{opt}}^{*(2)}]$ as a function of $1/n$ with an SEM error bars. The number of simulated instances is 1000.

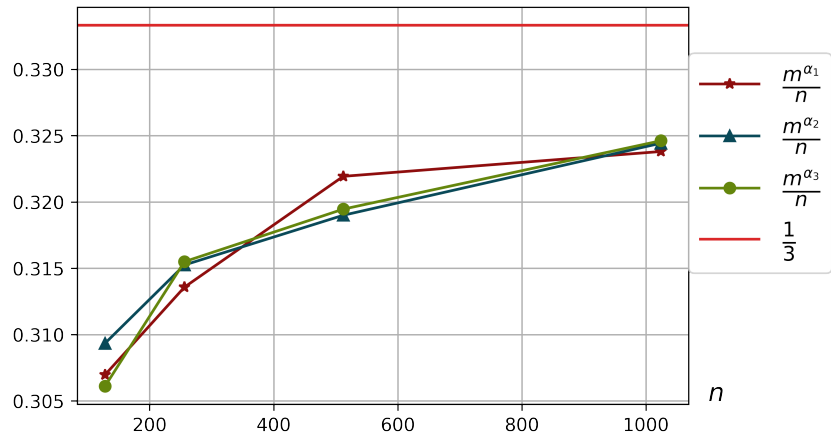


Figure B.5: The ratio of the number of connections in each direction to all when k is 3. The number of simulated instances is 500.

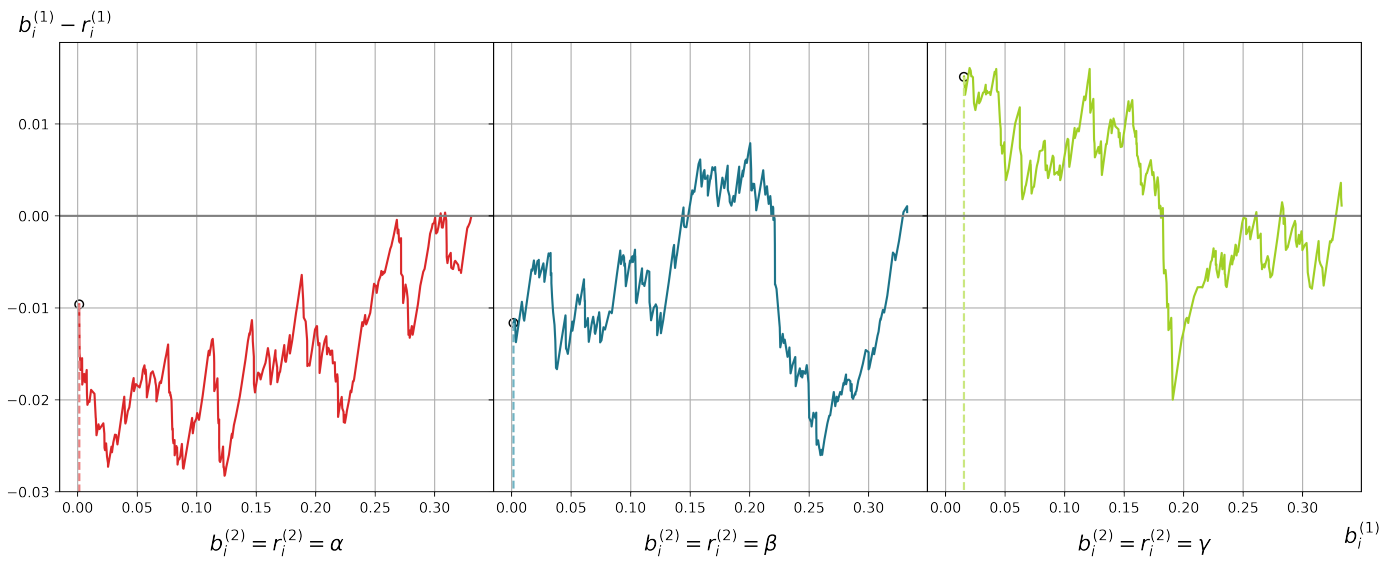


Figure B.6: Optimal transport field on each direction with $n = 1000$ and $k = 3$.

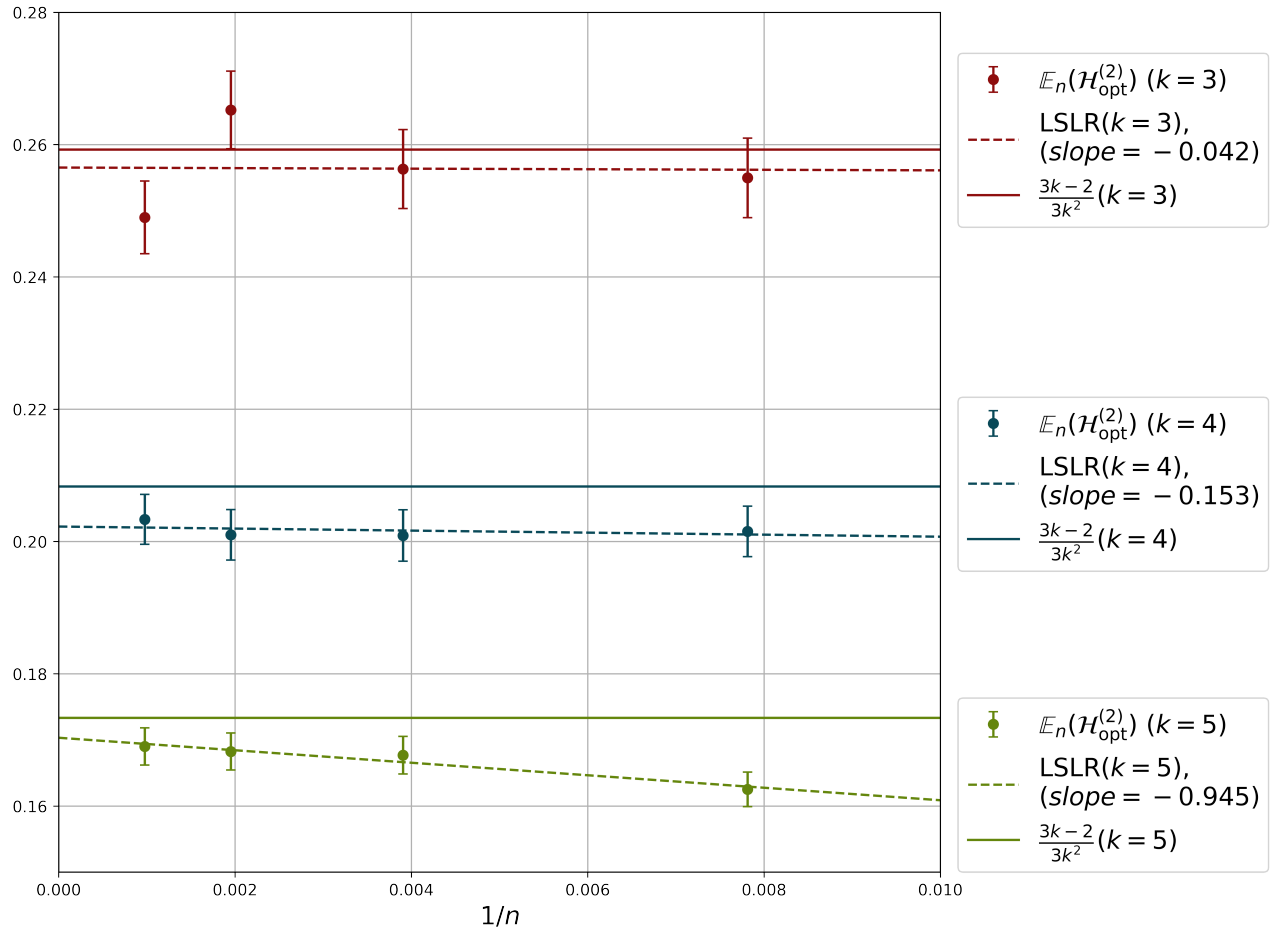


Figure B.7: Comparison between the numerical results and the results of equation (2). “LSLR” means “Least Squares Linear Regression”. The number of simulated instances is 1000.

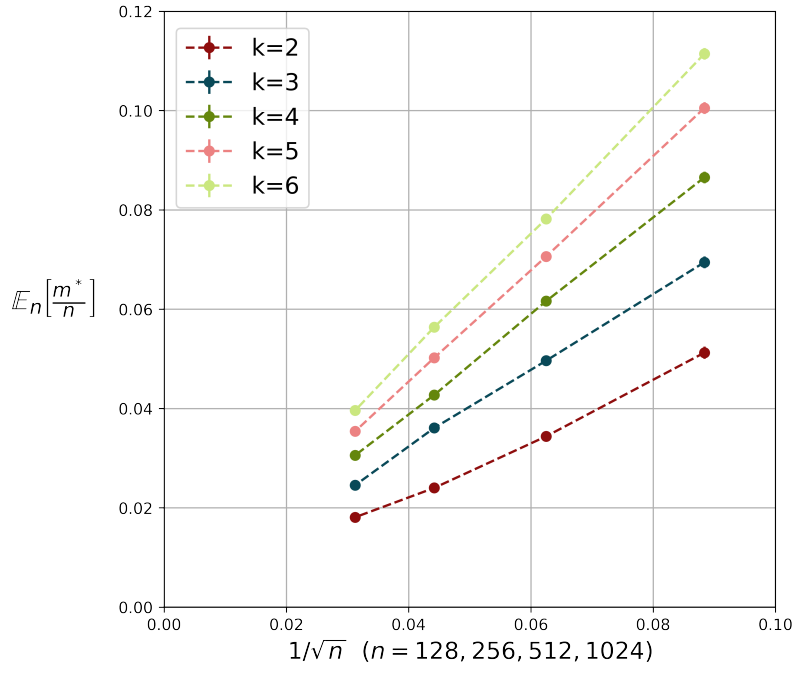


Figure B.8: $\mathbb{E}_n[\frac{m^*}{n}]$ as a function of $1/\sqrt{n}$ with an SEM error bars. The number of simulated instances is 1000.

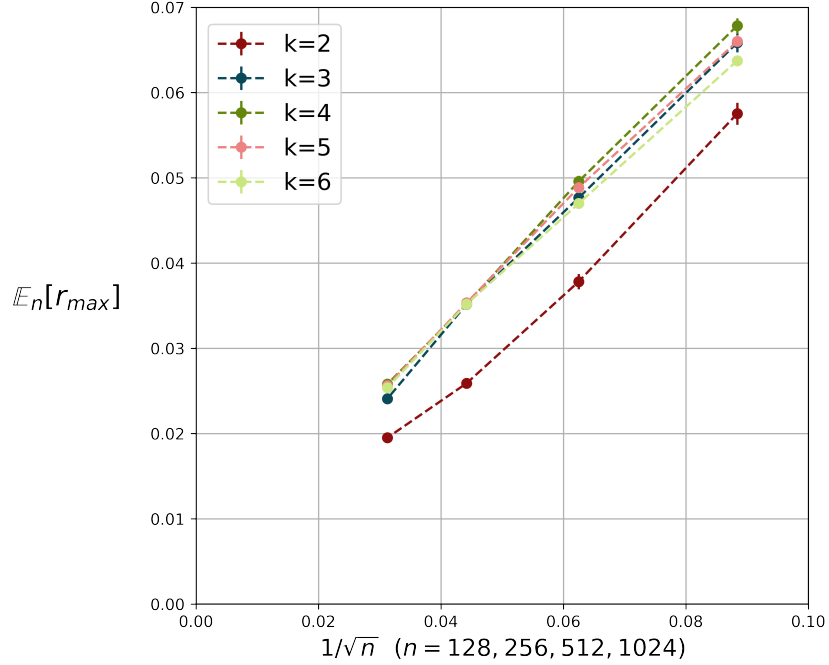


Figure B.9: $\mathbb{E}_n[r_{max}]$ as a function of $1/\sqrt{n}$ with an SEM error bars. The number of simulated instances is 1000.

References

- [1] Norman Biggs, E Keith Lloyd, and Robin J Wilson. *Graph Theory, 1736-1936*. Oxford University Press, 1986.
- [2] Gary Chartrand and Ping Zhang. *A first course in graph theory*. Courier Corporation, 2013.
- [3] Matteo D’Achille. *Statistical Properties of the Euclidean Random Assignment Problem*. PhD thesis, Université Paris-Saclay, 2020.
- [4] Marc Mézard and Giorgio Parisi. The euclidean matching problem. *Journal de Physique*, 49(12):2019–2025, 1988.
- [5] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.
- [6] Lev Landau and Evgeny Lifshitz. *Physique Théorique, Tome 5. Physique Statistique*.
- [7] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- [8] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.
- [9] Haim Brezis. Remarks on the monge–kantorovich problem in the discrete setting. *Comptes Rendus Mathématique*, 356(2):207–213, 2018.
- [10] Jim Pitman and Marc Yor. A guide to brownian motion and related stochastic processes. *arXiv preprint arXiv:1802.09679*, 2018.
- [11] Anatoly M Vershik. Kantorovich metric: Initial history and little-known applications. *Journal of Mathematical Sciences*, 133(4):1410–1417, 2006.
- [12] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [13] Elena Boniolo, Sergio Caracciolo, and Andrea Sportiello. Correlation function for the grid-poisson euclidean matching on a line and on a circle. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(11):P11023, 2014.
- [14] Sergio Caracciolo, Matteo D’Achille, and Gabriele Sicuro. Random euclidean matching problems in one dimension. *Physical Review E*, 96(4):042102, 2017.
- [15] Monroe D Donsker. Justification and extension of doob’s heuristic approach to the kolmogorov-smirnov theorems. *The Annals of mathematical statistics*, pages 277–281, 1952.

- [16] Wendelin Werner and Ellen Powell. Lecture notes on the gaussian free field. *arXiv preprint arXiv:2004.04720*, 2020.
- [17] Sergio Caracciolo and Gabriele Sicuro. One-dimensional euclidean matching problem: exact solutions, correlation functions, and universality. *Physical Review E*, 90(4):042112, 2014.
- [18] Sergio Caracciolo, Andrea Di Gioacchino, Enrico M Malatesta, and Luca G Molinari. Selberg integrals in 1d random euclidean optimization problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(6):063401, 2019.
- [19] Robert J McCann. Exact solutions to the transportation problem on the line. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455(1984):1341–1380, 1999.
- [20] Sergio Caracciolo, Matteo P D’Achille, Vittorio Erba, and Andrea Sportiello. The dyck bound in the concave 1-dimensional random assignment model. *Journal of Physics A: Mathematical and Theoretical*, 53(6):064001, 2020.
- [21] Sergio Caracciolo, Vittorio Erba, and Andrea Sportiello. The number of optimal matchings for euclidean assignment on the line. *Journal of Statistical Physics*, 183(3), 2021.
- [22] Roy Jonker and Anton Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325–340, 1987.
- [23] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
- [24] Leslie R Foulds. *Combinatorial optimization for undergraduates*. Springer Science & Business Media, 2012.