Random assignment problems for satellite data acquisition

Author:
Thomas Quarck

Supervisors:
Jessie Levillain
Matteo D'Achille

Contents

I Introduction **3**

I.1 Summary **3**

I.2 Assignment problem **3**
I.2.1 Assignment 3
I.2.2 Linear assignment 4
I.2.3 Euclidean random assignment problem (ERAP) 4

I.3 Typical euclidean greedy distance in $[0,1]^d$ **5**
I.3.1 Poisson process 5
I.3.2 Result and proof 5

II Theory **8**

II.1 A 1d model **8**
II.1.1 Setup 8
II.1.2 Mathematical model and ordering condition 8
II.1.3 Asymptotics of $\mathbb{P}(\pi_{opt} = id)$ (three regimes) 9

III Application to image acquisition **13**

III.1 Setup **13**

III.2 Quality of assignment **16**
III.2.1 Influence of parameters on the quality of assignment Q 16
III.2.1.1 Conclusions 17
III.2.2 1D model prediction 17

IV Perspectives **18**

IV.1 On the theoretic side **18**

IV.2 Application side **18**

Appendices **19**

A Hungarian Algorithm **19**
A.1 The algorithm 19
A.2 Complexity analysis 21
A.3 Test of the algorithm 22

A.3.1 PP-ERAP on $[0,1]$, $p=2$ 22

B Birkhoff polytope **27**
B.1 Birkhoff-von Neumann theorem 27
B.2 Birkhoff's Polytope 30

C Reminders **32**
C.1 Point processes 32
C.2 Poisson processes 34

References **37**

Part I
Introduction

I.1 Summary

This manuscript studies assignment problems with a focus on their stochastic formulation. We first review the linear assignment problem and a random variant called "Euclidean Random Assignment Problem", where cost are induced by distances. Then using a paper of Talagrand we derive a result on typical asymptotical distance of binomial processes. Next we introduced a simple one-dimensional toy model with a Poisson limit at the critical scale, inspired by a simulation of satellite images given the CNES. We then study the application of the Hungarian algorithm, to recover the original position of objects in these images. Particularly how the different parameters of the simulation impact the accuracy of the optimal assignment in this setup.

I.2 Assignment problem

I.2.1 Assignment

Sections I.2.1 and I.2.2 are largely inspired by [1].

The assignment problem asks how to assign $n \in \mathbb{N}$ items of one type (workers, sources) to n items of another type (tasks, targets). Mathematically, an assignment can be viewed as a bijection σ from a finite set U onto a finite set V both containing n elements. Identifying U and V with the set of integers $\{1, \dots, n\}$ allows us to represent an assignment by a permutation $\sigma \in \mathcal{S}_n$, where $\mathcal{S}_n$ denotes the symmetric group over a finite set of n symbols. We recall that every permutation is uniquely associated to an $n \times n$ matrix called a *permutation matrix* $P_\sigma = (p_{ij})_{i,j=1}^n$ defined by

$$p_{ij} = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

Equivalently, assignments can be represented via bipartite graphs. A graph $G = (U, V, E)$ with disjoint sets U and V and edge set E is *bipartite* if every edge connects a vertex of U with a vertex of V and no edge has both endpoints in only U or only V. A *matching* M in G is a subset of the edges such that every vertex of G meets at most one edge of the matching. Let's now suppose that $|U| = |V| = n$. In this case if every vertex of G is incident with an edge of the matching M, it is called a *perfect matching*. Every assignment can be represented as a perfect matching.

I.2.2 Linear assignment

We introduce a cost matrix $C = (c_{ij})_{ij=1}$ where c_{ij} is the cost of assigning i to j. We are looking for an assignment π_{opt} of minimal cost, i.e. minimizing the sum:

$$\mathcal{H}(\sigma) = \sum_{i=1}^{n} c_{i\sigma(j)} = \mathrm{Tr}(P_\sigma C). \tag{2}$$

By definition

$$\pi_{opt} = \arg\min_{\sigma} \sum_{i=1}^{n} c_{i\sigma(j)}. \tag{3}$$

And we call $\mathcal{H}_{opt}$ the minimal cost (or minimal energy)

$$\mathcal{H}_{opt} \coloneqq \mathcal{H}(\sigma_{opt}) = \min_{\sigma \in \mathcal{S}_n} \mathcal{H}(\sigma). \tag{4}$$

In 1955, Harold W. Kuhn found an algorithm called Hungarian algorithm (see [5]), Kuhn-Munkres algorithm which given a cost matrix finds the optimal assignment, in $O(n^3)$ time complexity. See appendix A for more information.

I.2.3 Euclidean random assignment problem (ERAP)

Whats follows comes from [3, Page 9-11].

Let $\mathcal{B} = \{B_i\}_{i=1}^n$ and $\mathcal{R} = \{R_i\}_{i=1}^n$ be two families of i.i.d. random variables distributed according to some measure ν, the disorder measure, on a Polish space $\mathcal{M}$ (typically $\mathbb{R}^d$).

The cost of assigning B_i to R_j is $c_{i,j} \coloneqq C(B_i, R_j)$ defined by some nonnegative measurable function $c : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$. The, $\{c_{ij}\}_{i,j=1}^n$ can be arranged into a non-symmetric, $n \times n$ matrix:

$$C = \begin{pmatrix} c(B_1, R_1) & c(B_2, R_2) & \cdots \\ \vdots & \ddots & \vdots \\ c(B_n, R_1) & \cdots & c(B_n, R_n) \end{pmatrix}$$

called the cost matrix. In our case, we will consider $\mathcal{M} = \mathbb{R}^d$ and the cost function c will be the euclidean distance $\mathcal{D}$ raised to the power p

$$c(x, y) = \mathcal{D}(x, y)^p, \quad p > 0. \tag{5}$$

The p comes from the p-Wasserstein distance from the continuous version of this problem. π_{opt} and $\mathcal{H}_{opt}$ are defined as in the previous section I.2.2, they are now random variables.

A well defined stochastic assignment problem, called ERAP (from the English "Euclidean Random Assignment Problem"), is specified by the tuple $(\mathcal{M}, \mathcal{D}, \nu, p)$.

I.3 Typical euclidean greedy distance in $[0,1]^d$

I.3.1 Poisson process

A Poisson point process (PPP) is a random countable set of points scattered locally compact metric space endowed with a Radon measure. The numbers of points in any disjoint regions are independent and, for a region A the number of points $N(A)$ follows a Poisson law of parameter given by the intensity measure of the process λ, of A. See appendix C for details.

I.3.2 Result and proof

Let η_1 be a Poisson process on the unit cube $[0,1]^d$, with intensity measure given by Leb, the Lebesgue measure. Equipped with the euclidean distance $\mathcal{D}$, $([0,1]^d, \mathcal{D})$ is a complete separable metric space. Thus η is a proper point process, which we denote by Π. We also define, for $A, B \subset [0,1]^d$,

$$D(A,B) = \min_{(a,b)\in A\times B} \mathcal{D}(a,b),$$

and use the shorthand $D(\{x\}, A) \coloneqq D(x,A)$.

Adapting an argument of the paper [9], we now prove that the average minimal distance between a typical point and a binomial point processes of sample size n and disorder Leb in the unit cube, has rate $n^{-1/d}$.

Proposition I.3.1. Let $n \in \mathbb{N}$. Conditioning the Poisson process on the event

$$\{\eta([0,1]^d) = n\},$$

We obtain a binomial process of sample size n and disorder Leb. In this setting and $n \to \infty$ for all $x \in [0,1]^d$,

$$K_d n^{-\frac{1}{d}} \leq \mathbb{E}[D(x,\Pi) \mid \eta([0,1]^d) = n] \leq 2K_d n^{-\frac{1}{d}} \tag{6}$$

Where $K_d \coloneqq \frac{\Gamma(\frac{1}{d})}{dC_d{}^{\frac{1}{d}}}$, and $C_d \coloneqq \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ is the volume of the unit ball in $\mathbb{R}^d$.

Proof. Upper Bound: for any $A \in \mathcal{M}$ and $x \in [0,1]^d$, let

$$\mathbb{P}_n(A) \coloneqq \mathbb{P}(A \mid \eta([0,1]^d) = n)$$

.

We first look at the tail of distribution of $D(x,\Pi_2)$ under $\mathbb{P}_n$,

$$\mathbb{P}_n\left(D(x,\Pi) > r\right) = \mathbb{P}_n(\eta(B(x,r)\cap[0,1]^d) = 0).$$

By Proposition C.1,

$$\mathbb{P}_n(\eta(B(x,r)\cap[0,1]^d) = 0) = \left(1 - \mathrm{Leb}\left(B(X,r)\cap[0,1]^d\right)\right)^n. \tag{7}$$

The volume of the intersection of a d-cube and a d-ball of fixed radius, whose center is constrained inside the d-cube is minimal when the center coincides with one of the corners of the d-cube. In this case, the total volume of the sphere is cut by a factor of $\frac{1}{2^d}$ (cut in half in all directions). The volume of the ball is proportional to r^d, we have the following inequality,

$$\text{Leb}\left(B(x,r) \cap [0,1]^d\right) \geq \frac{1}{2^d} C_d r^d. \tag{8}$$

Combining (7) and (8) we obtain,

$$\mathbb{P}_n\left(D(x,\Pi) > r\right) \leq \left(1 - \frac{1}{2^d} C_d r^d\right)^n.$$

but

$$\left(1 - \frac{1}{2^d} C_d r^d\right)^n \underset{n\to\infty}{\sim} e^{-n\frac{1}{2^d} C_d r^d}$$

Where $\sim$ refers to asymptotical equivalence[1]. For n large enough

$$\mathbb{P}_n\left(D(x,\Pi) > r\right) \leq e^{-n\frac{1}{2^d} C_d r^d} \tag{9}$$

The Euclidean distance between two points of the unit hypercube is at most $\sqrt{d}$. Giving us that, $\mathbb{P}_n(D(x,\Pi) > r) = 0$ for $r > \sqrt{d}$.

By integrating over r in (9),

$$\mathbb{E}_n[D(x,\Pi)] = \int_0^{\sqrt{d}} \mathbb{P}_n(D(x,\Pi) > r)\mathrm{d}r \leq \int_0^{\sqrt{d}} e^{-n\frac{1}{2^d} C_d r^d}\mathrm{d}r.$$

We now perform the asymptotic analysis of the integral $\int_0^{\sqrt{d}} e^{-n\frac{1}{2^d} C_d r^d}\mathrm{d}r$. Write $C'_d \coloneqq \frac{1}{2^d} C_d$. Performing the following change of variable:

$$\begin{cases} u = nC'_d r^d \\ \mathrm{d}u = dnC'_d r^{d-1}\mathrm{d}r \\ r = 0 \\ r = \sqrt{d} \end{cases} \longrightarrow \begin{cases} r = (\frac{u}{nC'_d})^{\frac{1}{d}} \\ \mathrm{d}r = \frac{u^{-\frac{d-1}{d}}}{d(nC'_d)^{\frac{1}{d}}}\mathrm{d}u \\ u = 0 \\ u = d^{\frac{d}{2}} nC'_d \end{cases}.$$

We get

[1] Given two sequences $(u_n)_{n\in\mathbb{N}}$, $(v_n)_{n\in\mathbb{N}}$ with $\forall n \in \mathbb{N}, v_n \neq 0$ we have that $\lim_{n\to\infty} \frac{u_n}{v_n} = 1$

$$\int_0^{\sqrt{d}} \exp\left(-nC'_d r^d\right) \mathrm{d}r = \int_0^{d^{\frac{d}{2}} nC'_d} \frac{u^{\frac{d-1}{d}}}{d(nC_d)^{\frac{1}{d}}} \exp\left(-u\right) \mathrm{d}u$$
$$= \frac{n^{-\frac{1}{d}}}{dC'^{\frac{1}{d}}_d} \int_0^{d^{\frac{d}{2}} nC'_d} u^{-\frac{d-1}{d}} \exp\left(-u\right) \mathrm{d}u$$
$$= n^{-\frac{1}{d}} C_n,$$

where $C_n = \frac{1}{dC'^{\frac{1}{d}}_d} \int_0^{d^{\frac{d}{2}} nC'_d} u^{-\frac{d-1}{d}} \exp\left(-u\right) du \underset{n\to\infty}{\longrightarrow} \frac{\Gamma(\frac{1}{d})}{dC'^{\frac{1}{d}}_d}$
We finally get the desired upper bound, for large n:

$$\mathbb{E}_n[D(x,\Pi)] \leq \frac{\Gamma(\frac{1}{d})}{dC'^{\frac{1}{d}}_d} n^{-\frac{1}{d}} = 2\frac{\Gamma(\frac{1}{d})}{dC_d^{\frac{1}{d}}} n^{-\frac{1}{d}} \tag{10}$$

Lower bound:

To prove the lower bound, we proceed similarly as what was done for the upper bound. Since

$$\mathrm{Leb}\left(B(x,r) \cap [0,1]^d\right) \leq \mathrm{Leb}\left(B(x,r)\right) = C_d r^d,$$

and

$$\mathbb{P}_n(D(x,\Pi) > r) = \mathbb{P}_n\left(\eta_2\left(B(x,r) \cap [0,1]^d\right) = 0\right) \tag{11}$$
$$= \left(1 - \mathrm{Leb}(B(x,r) \cap [0,1]^d)\right)^n \tag{12}$$
$$\geq \left(1 - C_d r^d\right)^n. \tag{13}$$

So using (13) we get by integrating over r that

$$\mathbb{E}_n[D(x,\Pi)] \geq \int_0^{\sqrt{d}} \left(1 - C_d r^d\right)^n \mathrm{d}r \sim \frac{\Gamma(\frac{1}{d})}{dC_d^{\frac{1}{d}}} n^{-\frac{1}{d}}. \tag{14}$$

□

Part II
Theory

II.1 A 1d model

II.1.1 Setup

The motivation for this internship was to apply assignment problem to satellite data acquisition. In particular recognition of objects on an image. Imagine we have n blue real points and n red shifted (see 4) points captured by the satellite. The general goal is to match the red points to their original blue point.

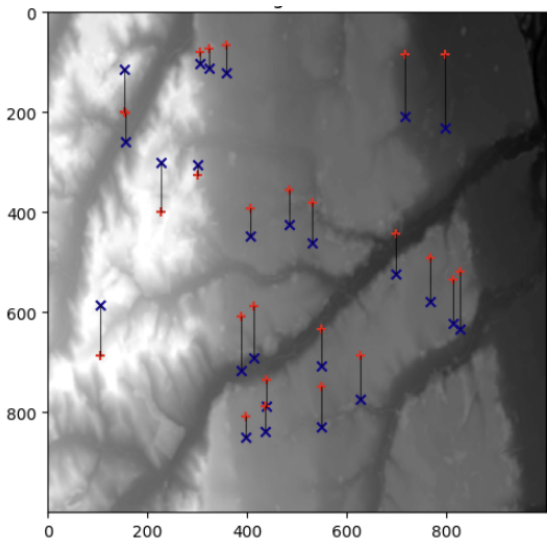


Figure 1: Optimal assignment for $n = 25$. In blue true points, in red points measured by the satellite with incidence angle $\theta_x = 0, \theta_y = 30$ for $h_{\text{bias}} = -100$

Seeing those simulations we noticed that the shift are all in roughly the same direction. That why we wondered if the problem was essentially unidirectional and tried to model the projection along the shift axis.

II.1.2 Mathematical model and ordering condition

Following I.2.3, let $(B_i)_{1\leq i\leq n}$ n blue uniform points on $[0,1]$ equipped with D^p, $p \geq 1$, $B_{(1)} < \ldots < B_{(n)}$ the ordered blue points (order statistics of the B_i's) and s_i be i.i.d. Rademacher random variables with $\mathbb{P}(s_i = \pm 1) = \frac{1}{2}$. We then define

$$R_i \coloneqq B_{(i)} + \varepsilon_n s_i. \tag{15}$$

For a convex cost ($p \geq 1$) , the monotonous assignment is optimal $\pi_{\text{opt}} \in \mathcal{S}_n$ (see section I.2.3) in one dimension is [3, Page 21]: the i-th red ordered point is matched to the i-th ordered blue point. Hence

$$\pi_{\text{opt}} = \text{id} \quad \Longleftrightarrow \quad R_1 < \cdots < R_n. \tag{16}$$

This is equivalent, for all $i = 1, \ldots, n-1$, to

$$R_i < R_{i+1}, \tag{17}$$

which can be rewritten as follows.

$$\Delta_i \coloneqq B_{(i+1)} - B_{(i)} > \varepsilon_n (s_{i+1} - s_i). \tag{18}$$

By a quick analysis

- If $s_i = s_{i+1}$, then (18) is satisfied;
- Same if $(s_i, s_{i+1}) = (-1, +1)$;
- The only "problematic" case is when $(s_i, s_{i+1}) = (+1, -1)$ which forces $\Delta_i > 2\varepsilon_n$.

Define the number of *potential violations*

$$V_n \coloneqq \sum_{i=1}^{n-1} \mathbf{1}\{(s_i, s_{i+1}) = (+1, -1)\}\, \mathbf{1}\{\Delta_i \leq 2\varepsilon_n\}. \tag{19}$$

We will note $S_i \coloneqq \{(s_i, s_{i+1}) = (+1, -1)\}$ and $G_i \coloneqq \{\Delta_i \leq 2\varepsilon_n\}$ for compactness.

Then $\pi_{\text{opt}} = \text{id}$ iff $V_n = 0$.

II.1.3 Asymptotics of $\mathbb{P}(\pi_{opt} = id)$ (three regimes)

Proposition II.1.1. Let $\alpha > 0$, as $n \to \infty$,

$$\mathbb{P}(\pi_{\text{opt}} = \text{id}) \longrightarrow \begin{cases} 0, & \varepsilon_n \gg n^{-2}, \\ e^{-\alpha/2}, & \varepsilon_n = \alpha\, n^{-2} \\ 1, & \varepsilon_n \ll n^{-2}. \end{cases}$$

Proof. The proof use the method of moments to show that V_n converges int distribution towards a Poisson random variable of parameter $\frac{\alpha}{2}$

$$\mathbb{P}(\pi_{\text{opt}} = \text{id}) = \mathbb{P}(V_n = 0) \tag{20}$$

Qualitative analysis

Let us first observe that

$$\mathbb{E}[V_n] = \sum_{i=1}^{n-1} \mathbb{P}\left(S_i \cap G_i\right). \tag{21}$$

Since the $(s_i)_{i=1}^n$ are independent of the Δ_i

$$\mathbb{P}\left(S_i \cap G_i\right) = \mathbb{P}\left(S_i\right)\mathbb{P}\left(G_i\right). \tag{22}$$

And (see [8] for the law of Δ_i)

$$\mathbb{P}\left(S_i\right) = \frac{1}{4}; \quad \mathbb{P}\left(G_i\right) = 1 - (1-2\varepsilon_n)^n. \tag{23}$$

Combining (21), (22) and (23) we get

$$\mathbb{E}[V_n] = \sum_{i=1}^{n-1} \frac{1-(1-2\varepsilon_n)^n}{4} = (n-1)\frac{1-(1-2\varepsilon_n)^n}{4}. \tag{24}$$

We can now compute an equivalent for $\mathbb{E}[V_n]$ as $n \to \infty$

$$\mathbb{E}[V_n] \sim (n-1)\frac{n\varepsilon_n}{2}. \tag{25}$$

We can now see why α/n^2 is the critical scale. If $\varepsilon_n << \alpha/n^2$, then $\mathbb{E}[V_n] \to 0$, and since $V_n \geq 0$, we have that almost surely $\lim V_n = 0$. Equivalently

$$\mathbb{P}(\pi_{\text{opt}} = \text{id}) = \mathbb{P}(V_n = 0) = 1. \tag{26}$$

On the other hand, if $\varepsilon_n >> \alpha/n^2$ then, * $\mathbb{E}[V_n]$ diverges. We cannot say anything rigorous with this statement alone but we have an infinite number of bad gaps on average making hard for $(\pi_{\text{opt}} = \text{id})$ to happen. Finally, the critical scale is the only case where $0 < \lim \mathbb{E}[V_n] < \infty$.

Higher order moments for critical scale $\varepsilon_n = \frac{\alpha}{n^2}$

We will now prove the following convergence in distribution as $n \to \infty$ $V_n \Rightarrow \text{Poisson}(\frac{\alpha}{2})$. In order to do so we will use the methods of moment, using factorial moments.

Indeed Lemma 3.13 from [4, page 14] states that, for $X = \sum_{i=1}^{N} Y_i$ a sum of indicators, Y_i might be dependent, the factorial moment of order k denoted $[X]_k$ is given by

$$[X]_k = k! \sum_{1 \leq i_1, \ldots, i_k \leq N} \mathbb{E}[Y_{i_1} \ldots Y_{i_k}],$$

We can directly apply this result to V_n:

$$[V_n]_k = k! \sum_{1 \leq i_1, \ldots, i_k \leq n-1} \mathbb{E}[X_{i_1} \ldots X_{i_k}].$$

Where the $X_i \coloneqq \mathbf{1}_{S_i}\mathbf{1}_{G_i}$ (see below (19)). Let us dive into the calculation of $E[X_{i_1}\dots X_{i_k}]$

$$E[X_{i_1}\dots X_{i_k}] = \mathbb{P}\left(\left[\bigcap_{j=1}^{k} S_{i_j}\right] \cap \left[\bigcap_{j=1}^{k} G_{i_j}\right]\right). \tag{27}$$

If for a choice of $\{i_1,\dots,i_k\}$ there are two indices i_j, i_m such that $|i_j - i_m| = 1$ (two successive indices), then $\bigcap_{j=1}^{k} S_{i_j} = \emptyset$, and hence in this case $E[X_{i_1}\dots X_{i_k}] = 0$. Let us now consider the choices of $\{i_1,\dots,i_k\} \in I_k$ where $I_k \coloneqq \{\{i_1,\dots,i_k\} \mid \forall i_j, i_m, |i_j - i_m| \geq 2\}$ (no consecutive indices). In this case all, the S_{i_j} are independent for $j = 1,\dots,k$ and also independent of any of the G_i. Thus

$$\mathbb{P}\left(\left[\bigcap_{j=1}^{k} S_{i_j}\right] \cap \left[\bigcap_{j=1}^{k} G_{i_j}\right]\right) = \prod_{j=1}^{k} \mathbb{P}\left(S_{i_j}\right) \mathbb{P}\left(\bigcap_{j=1}^{k} G_{i_j}\right) = \left(\frac{1}{4}\right)^k \mathbb{P}\left(\bigcap_{j=1}^{k} G_{i_j}\right).$$

Therefore

$$E[X_{i_1}\dots X_{i_k}] = \left(\frac{1}{4}\right)^k \mathbb{P}\left(\bigcap_{j=1}^{k} G_{i_j}\right). \tag{28}$$

We now have to compute $\mathbb{P}\left(\bigcap_{j=1}^{k} G_{i_j}\right)$. Since the Δ_i are exchangeable, $\mathbb{P}\left(\bigcap_{j=1}^{k} G_{i_j}\right) = \mathbb{P}\left(\bigcap_{j=1}^{k} G_j\right)$. Then, by the inclusion-exclusion principle applied to the complementary event, we get

$$\mathbb{P}\left(\bigcup_{j=1}^{k} G_j^c\right) = \sum_{j=1}^{k}(-1)^{j-1} \sum_{1\leq a_1<\dots<a_j\leq k} \mathbb{P}\left(\bigcap_{i=1}^{j} G_{a_i}^c\right). \tag{29}$$

Using now the exchangeability of the Δ_i, $\mathbb{P}\left(\bigcap_{i=1}^{j} G_{a_i}^c\right) = \mathbb{P}\left(\bigcap_{i=1}^{j} G_i^c\right)$ we get

$$\mathbb{P}\left(\bigcap_{i=1}^{j} G_i^c\right) = \int_{\mathbb{R}^{n+1}} \mathbf{1}_{\bigcap_{i=1}^{j} x_i > 2\varepsilon_n} f_\Delta(x)dx, \tag{30}$$

where $f_\Delta(x) = n!\mathbf{1}\{\sum_{i=1}^{n+1} x_i = 1; x_i \geq 0, \forall i\}$ is the density function of the vector $\Delta = (\Delta_1,\dots,\Delta_{n+1})$. We then perform the following change of variables

$$y_i = \begin{cases} x_i - 2\varepsilon_n & \text{if} \quad 1 \leq i \leq j \\ x_i & \text{otherwise} \end{cases},$$

whose absolute value of the determinant of the Jacobian is equal to 1. We have

$$\int_{\mathbb{R}^{n+1}} \mathbf{1}_{\bigcap_{i=1}^{j} x_i > 2\varepsilon_n} n!\mathbf{1}\{\sum_{i=1}^{n+1} x_i = 1; x_i \geq 0, \forall i\}dx = \int_{\mathbb{R}^{n+1}} n!\mathbf{1}\{\sum_{i=1}^{n+1} y_i = 1-2j\varepsilon_n; y_i \geq 0, \forall i\}dy$$

We now perform a second change of variables, for all $i = 1, \ldots, n+1$,

$$z_i = \frac{y_i}{1-2j\varepsilon_n},$$

whose Jacobian is $(1-2j\varepsilon_n)^{n+1}$. So

$$\begin{aligned}\int_{\mathbb{R}^{n+1}} n!\mathbf{1}\{\sum_{i=1}^{n+1} y_i = 1-2j\varepsilon_n; y_i \geq 0, \forall i\}dy &= \int_{\mathbb{R}^{n+1}} n!\mathbf{1}\{\sum_{i=1}^{n+1} z_i = 1; z_i \geq 0, \forall i\}(1-2j\varepsilon_n)^{n+1}dz \\ &= (1-2j\varepsilon_n)^{n+1}\int_{\mathbb{R}^{n+1}} f_\Delta(z)dz \\ &= (1-2j\varepsilon_n)^{n+1}.\end{aligned}$$

Coming back to (29), we finally get,

$$\mathbb{P}\left(\bigcup_{j=1}^{k} G_j^c\right) = \sum_{j=1}^{k}(-1)^{j-1}\sum_{1\leq a_1<\ldots<a_j\leq k}(1-2j\varepsilon_n)^{n+1}. \tag{31}$$

But $\{1 \leq a_1 < \ldots < a_j \leq k\}$ has $\binom{k}{j}$ terms. Injecting (31) into (28) we obtain

$$E[X_{i_1}\ldots X_{i_k}] = \begin{cases} \left(\frac{1}{4}\right)^k\left(1-\sum_{j=1}^{k}(-1)^{j-1}\binom{k}{j}(1-2j\varepsilon_n)^{n+1}\right) & \text{if } \{i_1,\ldots,i_k\} \in I_k \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

Therefore the k-th factorial moment of V_n writes exactly as

$$[V_n]_k = k!|I_k|\left(\frac{1}{4}\right)^k\left(1-\sum_{j=1}^{k}(-1)^{j-1}\binom{k}{j}(1-2j\varepsilon_n)^{n+1}\right).$$

But $|I_k| = \binom{n-k}{k}$. Indeed we have to choose k numbers from $\{1,\ldots,n\}$ with at least one gap between any two picks. Think of reserving the $k-1$ gaps to separate them. That leaves $(n-1)-(k-1) = n-k$ usable positions. Now choose which k of these $n-k$ giving $\binom{n-k}{k}$. We finally obtain the following exact formula

$$[V_n]_k = k!\binom{n-k}{k}\left(\frac{1}{4}\right)^k\left(1-\sum_{j=1}^{k}(-1)^{j-1}\binom{k}{j}(1-2j\varepsilon_n)^{n+1}\right)$$

We now have to compute the limit as $n \to \infty$ of $[V_n]_k$. Recall that

$$(1-2j\varepsilon_n)^{n+1} = e^{-2j\frac{\alpha}{n}+o(1/n)}.$$

So

$$\begin{aligned}\sum_{j=1}^{k}(-1)^{j-1}\binom{k}{j}(1-2j\varepsilon_n)^{n+1} &= 1-\left(1-e^{-2\frac{\alpha}{n}+o(1/n)}\right)^k \\ &= 1-\left(2\frac{\alpha}{n}+o(1/n)\right)^k\end{aligned}$$

moreover

$$k!\binom{n-k}{k} \sim n^k,$$

combining all

$$[V_n]_k \sim n^k\left(\frac{1}{4}\right)^k\left(2\frac{\alpha}{n}\right)^k \to \left(\frac{\alpha}{2}\right)^k.$$

[Thomas: super critical case is missing] □

Part III
Application to image acquisition

III.1 Setup

In our simulated data, for each true point (blue), the simulator generates several projected locations (red) corresponding to different height bias. Indeed the red points are obtained using the following equation,

$$(x_r, y_r) = (x_b, x_b) + \Delta h(\tan(\theta_x), \tan(\theta_y)), \tag{33}$$

with

$$\Delta h = \text{topography}(x_b, x_b) + h_{\text{bias}}. \tag{34}$$

In our simulated data the topography is fixed (see figure below).

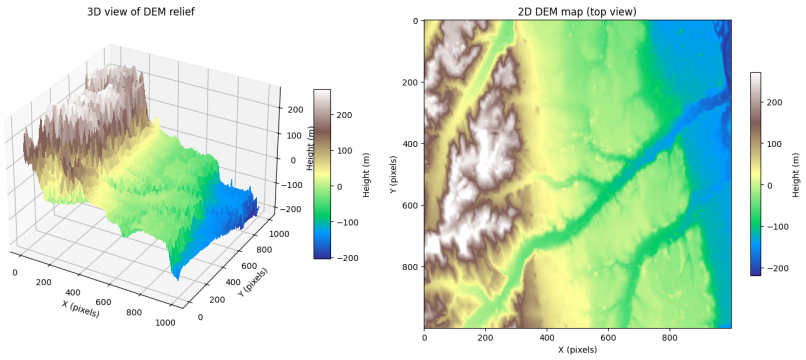


Figure 2: Topography of our simulation 2 pixels = 1 meter

And the height distribution is the following:

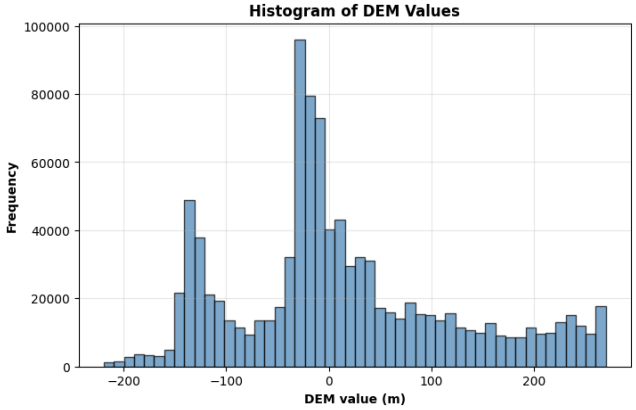


Figure 3: Height distribution

The satellite is sufficiently far away to suppose any variation in theta for different objects negligible.

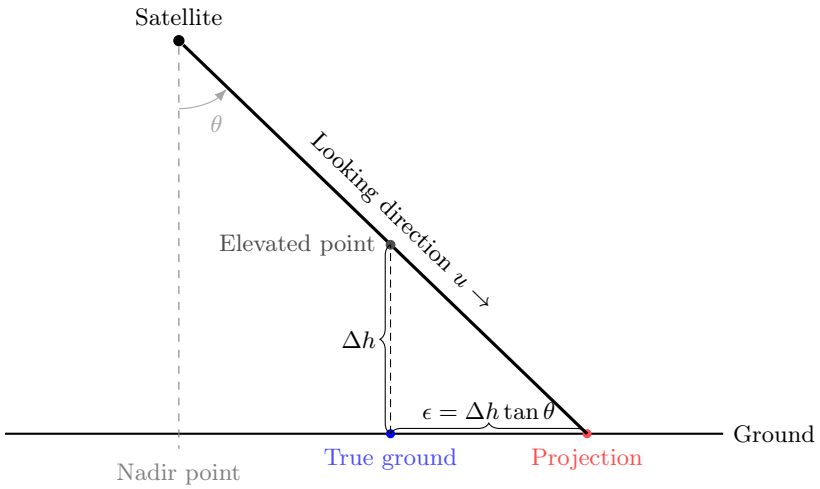


Figure 4: Geometry of satellite imaging. An elevated object of height Δh is shifted by $\epsilon = \Delta h \tan\theta$ along the look direction u; θ is the incidence angle from nadir.

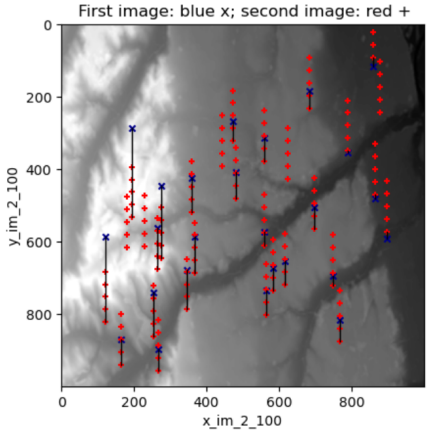


Figure 5: In blue true points, in red points measured by the satellite with incidence angle $\theta_x = 0, \theta_y = 30$ for $h_{\text{bias}} \in [-100, -50, 0, 50, 100]$

Last parameter to consider in this simulation is the number of missing true points (blue points) n_{missing}.

To sum up parameters of the problem are

- the number of objects n,
- n_{missing},

- angle of incidence θ_y (and θ_x which we can fix equal to zero since it only changes the direction the projection).
- h_{bias}.

III.2 Quality of assignment

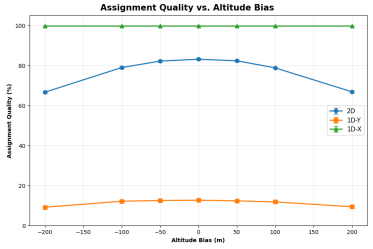
We want to analyze the impact of the different quantities of our simulation on the quality of the optimal assignment. To do so we compare it to the true assignation via the ratio:

$$Q_{\text{empirical}} = \frac{\#\{\text{correctly assigned points}\}}{\#\{\text{blue points}\}}.$$

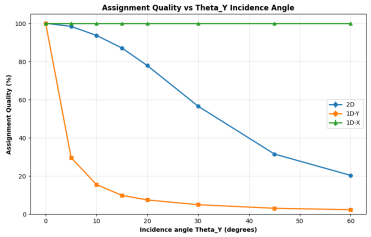
Recall that $\#\{\text{blue points}\} = n - n_{\text{missing}}$. We will study the ratio in the 2D case and its 1D projection (x and y directions).

III.2.1 Influence of parameters on the quality of assignment Q

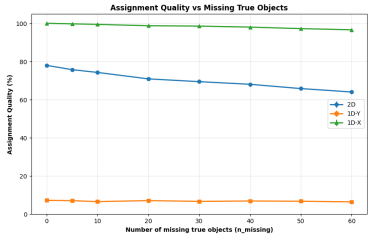
In this section we will study empirically the impact of parameter on Q. Here is the plot of Q in function of the parameters, for the 2D and 1D assignments.



(a) Impact of h_{bias}. Parameters: n=50, n_{iter}=1000, n_{missing}=5, θ_y=20°.



(b) Impact of θ_y. n=100, n_{missing}=0, h_{bias}=0, n_{iter}=300.



(c) Impact of n_{missing}. n=100, θ_y=20°, h_{bias}=0, n_{iter}=300

(d) Impact of n. h_{bias}=0, n_{iter}=300, n_{missing}=10, θ_y=20°.

Figure 6: Impact of each parameter on Q, error=2*SE.

From those plots we observe :

- An increase in θ_y, n_{missing}, or n leads to a decrease of Q for the 2D and 1D Y assignments.
- For h_{bias} we observe the same behavior but for an increase in its absolute value, explaining the parabola.
- Given the model on the projection perpendicular to the shift axis, projections of blue and red points are the same. This explains why we observe 100% accuracy.

III.2.1.1 Conclusions

In this setup the best use of Hungarian algorithm, seem to be on the projection along the perpendicular axis of the shift. In the 2D case it performs reasonably well for low values of the parameters, but never beats the perpendicular projection. Lastly the collinear projection performs poorly in every case.

III.2.2 1D model prediction

Our analysis reveals that projecting along the x-axis is the best to get maximum accuracy. To make it more interesting we can introduce some 'noise' along the x-axis. Placing in the setup of our model, a Rademacher of parameter $\frac{1}{2}$, scaled by $\varepsilon > 0$. We recall that as shown in the proof of II.1.1 the expected number of violations is

$$\mathbb{E}[V_n] = (n-1)\frac{1-(1-2\varepsilon)^n}{4}$$

We can give a very elementary lower bound on $\mathbb{P}(\pi_{\text{opt}} = \text{id})$ using Markov's inequality,

$$\mathbb{P}(V_n = 0) \geq 1 - \mathbb{E}[V_n]$$

Part IV
Perspectives

IV.1 On the theoretic side

- Use a richer noise. Replace the $\varepsilon_n s_i$ Rademacher per continuous noise (uniform, gaussian)
- Link the model to the rectangular case n reds, $n - n_{\text{missing}}$ blue
- Go beyond one dimension, using Fourrier analysis in two dimension in the case $p = 2$ to get estimates on the cost of $\mathbb{V}(\mathcal{H}_{\text{opt}})$.
- Prove the super critical scale of the proposition II.1.1

IV.2 Application side

- Broaden the scope of the tests performed
- Use other simulations/real data
- Comparison of Hungarian vs Greedy algorithm
- In the setup subsection I.3.2, compare as n grows the optimal assignment, to the accuracy of the optimal assignment looking only a points at a distance of the order of closest neighbor $n^{-1/d}$. This equivalent as using the Hungarian algorithm on the truncated (sparser) distance matrix.

A Hungarian Algorithm

A.1 The algorithm

The Hungarian algorithm, also known as the Kuhn-Munkres algorithm, is used to solve assignment problems by finding an optimal assignment that minimizes the total cost of a given cost matrix. See the original article [5]. My implementation of the code is available here: github.com/ThomasQuarck/hungarian-assignment.

Duality Condition

In the context of the assignment problem, duality refers to the use of dual variables (or potentials) u_i and v_j associated with each row i and column j, respectively. The duality condition helps achieve optimality. Specifically, for a given cost matrix C with elements c_{ij}, the dual variables satisfy the following condition for all i and j:

$$u_i + v_j \leq c_{ij}.$$

Equality is achieved when an assignment is made from row i to column j. Thus the complete assignment is obtained when we have n of theses equalities. The algorithm iteratively adjusts these dual variables to ensure that an optimal assignment can be obtained.

Steps of the Algorithm

- **Initialization**:

  Start with the original cost matrix. Initialize dual variables $u_i = 0$ for all rows and $v_j = 0$ for all columns.

- **Subtract Minimum in Rows and Columns**:

  For each row and column, subtract the minimum value to ensure there is at least one zero in each row and column.

- **Covering Zeros**:

  Cover all zeros in the matrix with a minimum number of horizontal and vertical lines. If the number of lines equals the size of the matrix, an optimal assignment exists among the zeros. Otherwise, proceed to the adjustment step.

- **Adjusting the Matrix**:

  Determine the smallest non-covered element, denoted as δ. Subtract δ from all non-covered elements and add it to elements that are covered by two lines (intersection points). Return to the covering step.

Example

We start with the initial cost matrix where each element represents the cost of assigning a worker to a task:

$$\begin{pmatrix} 5 & 3 & 8 & 6 \\ 4 & 6 & 5 & 7 \\ 7 & 2 & 9 & 8 \\ 6 & 4 & 7 & 5 \end{pmatrix}$$

Initial Potential Vectors

The potential vectors u and v are initialized to zero:

$$u = [0, 0, 0, 0], \quad v = [0, 0, 0, 0]$$

Adjusting the Matrix by Subtracting Row Minima

We find the minimum value in each row and subtract it from all elements in that row:

$$\begin{pmatrix} 2 & 0 & 5 & 3 \\ 0 & 2 & 1 & 3 \\ 5 & 0 & 7 & 6 \\ 2 & 0 & 3 & 1 \end{pmatrix}$$

$$u = [3, 4, 2, 4] \quad v = [0, 0, 0, 0]$$

Adjusting the Matrix by Subtracting Column Minima

Next, we find the minimum value in each column and subtract it from all elements in that column:

$$\begin{pmatrix} 2 & 0 & 4 & 2 \\ 0 & 2 & 0 & 2 \\ 5 & 0 & 6 & 5 \\ 2 & 0 & 2 & 0 \end{pmatrix}$$

$$u = [3, 4, 2, 4] \quad v = [0, 0, 1, 1]$$

Covering of Zeros and Finding δ We observe that only 3 lines (rows 2 and 4 and column 2) are needed to cover all the zeros. We identify the smallest non-covered element, which is 2. This is our δ:

$$\delta = \min\{\text{non-covered elements}\} = 2$$

Adjusting the Matrix with δ

We adjust the matrix by subtracting δ from non-covered elements and adding it to elements covered by two lines:

$$\begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 4 & 0 & 2 \\ 3 & 0 & 4 & 3 \\ 2 & 2 & 2 & 0 \end{pmatrix}$$

$$u = [5, 4, 4, 4] \quad v = [0, -2, 1, 1]$$

In this matrix, we box the independent zeros and cover the necessary lines:

$$\begin{pmatrix} \boxed{0} & 0 & 2 & 0 \\ 0 & 4 & \boxed{0} & 2 \\ 3 & \boxed{0} & 4 & 3 \\ 2 & 2 & 2 & \boxed{0} \end{pmatrix}$$

Lines covering all zeros are rows 1 and 3 and column 3.

The algorithm continues iteratively until an optimal assignment is found, which is when each row and each column contain exactly one covered zero. In the final matrix, this condition will be met, and the optimal assignment can be determined by selecting these unique covered zeros.

A.2 Complexity analysis

This section summarize the complexity analysis of my implementation of the hungarian algorithm and compare it to the Scipy version

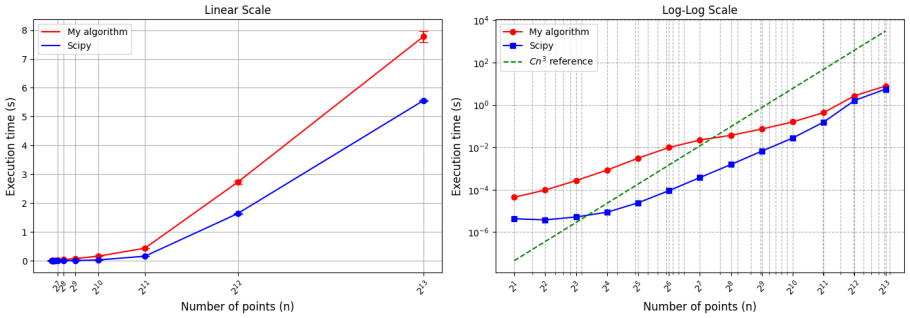


Figure 7: Complexity analysis on matrix of sizes $n = 2^j$ with j in $[|2, 13|]$ (1000 realizations per matrix size).

Table 1: Performance comparison between our implementation and Scipy

Size	Algorithm	Mean (s)	Standard Deviation	95% CI	Ratio (Scipy/Ours)
2	Ours	0.000 044 26	0.000 021 28	0.000 001 35	0.096 613
2	Scipy	0.000 004 28	0.000 042 27	0.000 002 67	
4	Ours	0.000 096 59	0.000 060 42	0.000 003 82	0.039 100
4	Scipy	0.000 003 78	0.000 008 43	0.000 000 53	
8	Ours	0.000 269 68	0.000 140 16	0.000 008 86	0.019 187
8	Scipy	0.000 005 17	0.000 012 65	0.000 000 80	
16	Ours	0.000 841 96	0.000 344 76	0.000 021 80	0.010 319
16	Scipy	0.000 008 69	0.000 009 39	0.000 000 59	
32	Ours	0.003 072 20	0.000 942 49	0.000 059 61	0.007 808
32	Scipy	0.000 023 99	0.000 010 89	0.000 000 69	
64	Ours	0.009 817 52	0.002 468 84	0.000 156 14	0.009 214
64	Scipy	0.000 090 46	0.000 014 29	0.000 000 90	
128	Ours	0.022 074 38	0.004 936 24	0.000 312 20	0.016 933
128	Scipy	0.000 373 79	0.000 045 12	0.000 002 85	
256	Ours	0.036 781 22	0.008 010 85	0.000 506 65	0.041 800
256	Scipy	0.001 537 44	0.000 158 21	0.000 010 01	
512	Ours	0.072 718 09	0.013 695 47	0.000 866 18	0.090 805
512	Scipy	0.006 603 18	0.000 518 32	0.000 032 78	
1024	Ours	0.158 074 95	0.025 520 31	0.001 614 05	0.174 307
1024	Scipy	0.027 553 60	0.002 028 98	0.000 128 32	

A.3 Test of the algorithm

We perform here some checks of the algorithm using known results.

A.3.1 PP-ERAP on $[0,1]$, $p=2$

Let us consider a PP-ERAP on $[0,1]$ and $p=2$. According to [3, page 27], we know that for $p>1$ in the one dimensional case

$$\mathbb{E}[\mathcal{H}_{opt}] \underset{n\to\infty}{\sim} c_p n^{1-\frac{p}{2}}. \tag{35}$$

In the case $p=2$, we should get $\mathbb{E}[\mathcal{H}_{opt}] \underset{n\to\infty}{\longrightarrow} c_2$.

Proposition A.1. The average moment of order 2 of the Brownian bridge on $[0,1]$ is c_2.

Since $p>1$ we are in the convex regime thus the solution must be ordered. We can then write,

$$\mathcal{H}_{opt} = \sum_{i=1}^{n} (B_{(i)} - R_{(i)})^2. \tag{36}$$

We write the empirical CDF of $\mathcal{B}$ and $\mathcal{R}$

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{B_i \leq x\}}, \quad G_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{R_i \leq x\}}.$$

If $t_i = \frac{i}{n+1}$ then,

$$B_{(i)} = F_n^{-1}(t_i), \quad R_{(i)} = G_n^{-1}(t_i).$$

This allows use to rewrite,

$$\mathcal{H}_{opt} = \sum_{i=1}^{n} (F_n^{-1}(t_i) - G_n^{-1}(t_i))^2. \tag{37}$$

But,

$$\frac{1}{n}\sum_{i=1}^{n} (\sqrt{n}F_n^{-1}(t_i) - \sqrt{n}G_n^{-1}(t_i))^2 = \int_0^1 (\sqrt{n}F_n^{-1}(t) - \sqrt{n}G_n^{-1}(t))^2 dt + o(1). \tag{38}$$

Recall that $(B_i)_{i=1}^n$ and $(R_i)_{i=1}^n$ are independent i.i.d uniform variable of CDF $F(x) = G(x) = x$. So $F^{-1}(x) = G^{-1}(x) = x$. Using this in (38) combined with (37), we get,

$$\mathcal{H}_{opt} = \int_0^1 (\sqrt{n}(F_n^{-1}(t) - F^{-1}(t)) - \sqrt{n}(G_n^{-1}(t) - G^{-1}(t))^2 dt. \tag{39}$$

By Donsker theorem, $\sqrt{n}(F_n^{-1}(t) - F^{-1}(t)) \Rightarrow -B^{(1)}(t)$ and $\sqrt{n}(G_n^{-1}(t) - G^{-1}(t)) \Rightarrow -B^{(2)}(t)$, where $B^{(2)}$ and $B^{(1)}$ are two independent Brownian bridges. Finally

$$\mathcal{H}_{opt} \Rightarrow \int_0^1 (B^{(2)}(t) - B^{(1)}(t))^2 dt. \tag{40}$$

By taking expectations, and by independence of $B^{(2)}$, $B^{(1)}$,

$$\mathbb{E}[\mathcal{H}_{opt}] = \int_0^1 \mathbb{E}[(B^{(2)}(t) - B^{(1)}(t))^2] dt = 2\int_0^1 t(1-t) dt = \frac{1}{3}. \tag{41}$$

In practice we observe the following behavior for $\mathbb{E}[\mathcal{H}_{opt}]$.

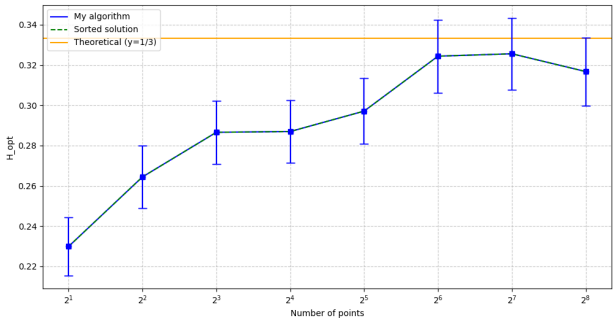


Figure 8: Empirical mean of H_{opt} for n points in $[0,1]$ $n = 2, 4, 8, 16, 32, 64, 128, 256$ (1000 realizations per number of points).

Indeed, if we define the optimal transport field [3, page 24] as:

$$\mu_i = R_{\pi_{opt}(i)} - B_i.$$

We can then write $\mathcal{H}_{opt}$ in terms of μ as follows for any p

$$\mathcal{H}_{opt} = \sum_{i=1}^{n} |\mu_i|^p. \tag{42}$$

We can now compute $\mathbb{E}[\mathcal{H}_{opt}]$ in the case $p = 2$

$$\mathbb{E}[\mathcal{H}_{opt}] = \mathbb{E}\left[\sum_{i=1}^{n}(R_{\pi_{opt}(i)} - B_i)^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[R^2_{\pi_{opt}(i)}\right] - 2\sum_{i=1}^{n}\mathbb{E}\left[R_{\pi_{opt}(i)}B_i\right] + \sum_{i=1}^{n}\mathbb{E}\left[B_i^2\right].$$

The B_i and R_i have the same law so for all i $\mathbb{E}\left[B_i^2\right] = \mathbb{E}\left[R^2_{\pi_{opt}(i)}\right]$

$$\mathbb{E}[\mathcal{H}_{opt}] = 2\sum_{i=1}^{n}\mathbb{E}\left[B_i^2\right] - 2\sum_{i=1}^{n}\mathbb{E}\left[R_{\pi_{opt}(i)}B_i\right].$$

We know that the solution is ordered meaning $R_{(i)}$ is send on the i^{th} blue $B_{(i)}$ (i^{th} order statistic). In the end using the independence of B_i and R_i

$$\mathbb{E}[\mathcal{H}_{opt}] = 2\sum_{i=1}^{n}\mathbb{E}\left[B_i^2\right] - 2\sum_{i=1}^{n}\mathbb{E}\left[B_{(i)}\right]^2.$$

We now have to compute $\mathbb{E}\left[B_i^2\right]$ and $\mathbb{E}\left[B_{(i)}\right]$.
As $B_i \sim \text{Uniform}([0,1])$

$$\mathbb{E}\left[B_i^2\right] = \int_{[0,1]} x^2 dx = \frac{1}{3},$$

and $B_{(i)} \sim \text{Beta}(i, n-i+1)$ (easily seen via the cdf) so

$$\mathbb{E}\left[B_{(i)}\right] = \frac{i}{n+1}.$$

Finally,

$$\mathbb{E}[\mathcal{H}_{opt}] = 2\left[\frac{n}{3} - \frac{n(2n+1)}{6(n+1)}\right] = \frac{n}{3(n+1)}. \tag{43}$$

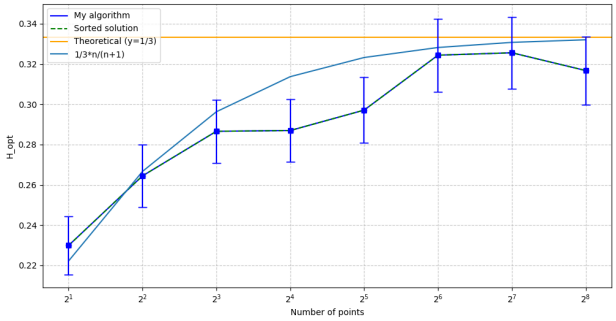


Figure 9: Comparaison of the empirical average of $\mathcal{H}_{\text{opt}}$ (blue dots with error bars) with the exact formula (43) (light blue continuous line). The limit value is plotted as an orange horizontal line. $n = 2, 4, 8, 16, 32, 64, 128, 256$ (1000 realizations per number of points)

Table 2: Empirical measures of $\mathcal{H}_{\text{opt}}$

Matrix Size	Mean Cost
2	0.2299 ± 0.0145
4	0.2645 ± 0.0154
8	0.2866 ± 0.0158
16	0.2870 ± 0.0157
32	0.2971 ± 0.0162
64	0.3244 ± 0.0181
128	0.3256 ± 0.0177
256	0.3168 ± 0.0170

Note: 95% confidence intervals shown

Rectangular case - exponential entries

For an $m \times n$ rectangular matrix with iid exponential entries of parameter 1 we know that [7, page 1] the expected value of the cost of optimal assignment is given exactly by:

$$\mathbb{E}[\mathcal{H}_{\text{opt}}] = \sum_{\substack{i,j\geq 0 \\ i+j<k}} \frac{1}{(m-i)(n-j)}. \tag{44}$$

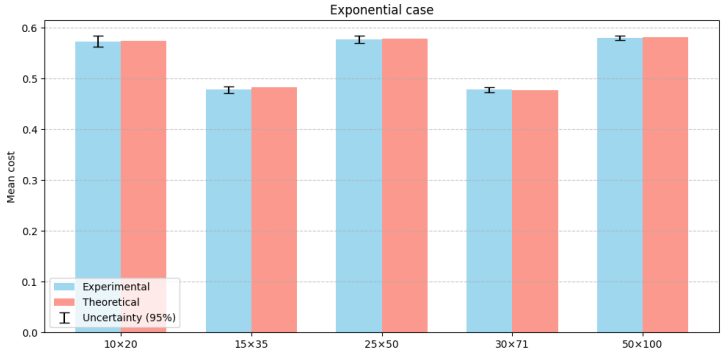


Figure 10: Mean cost for rectangular matrix of size: 10×20, 15×35, 25×50, 30×71, 50×100.

Table 3: Statistical validation of mean costs

Matrix Size	Experimental Cost	Theoretical Cost	Deviation (%)
10×20	0.573 ± 0.011	0.574	−0.2
15×35	0.478 ± 0.007	0.483	−0.9
25×50	0.577 ± 0.007	0.579	−0.3
30×71	0.478 ± 0.005	0.477	0.3
50×100	0.580 ± 0.005	0.581	−0.1

Note: 95% confidence intervals (k=2 expansion factor)

B Birkhoff polytope

B.1 Birkhoff-von Neumann theorem

In this section we will recall a proof from [2] of the Birkhoff-von Neumann theorem, using linear algebra.

Definition B.1. Permanent [2, Page 504-506]
If $A = (a_{i,j})_{1 \leq i,j \leq n}$ is an $n \times n$ matrix with complex entries the *permanent* of A, $\mathrm{per}(A)$ is defined by

$$\mathrm{per}(A) = \sum_{\sigma \in \mathcal{S}_n} a_{1,\sigma(1)} \dots a_{n,\sigma(n)}$$

Remark. Another formula [2, Page 504-506]

There is also a formula to compute $\mathrm{per}(A)$ that is analogous to computing the determinant by expanding into its minors.

$$\forall 1 \leq i \leq n, \mathrm{per}(A) = \sum_{j=1}^{n} a_{i,j} \mathrm{per}(\Delta_{i,j}). \tag{45}$$

Before proving the theorem, we will need two Lemmas.

Lemma B.1. [2, Page 504-506] Let $A \in \mathbb{R}^{n \times n}$ be a non-negative matrix. Then $\mathrm{per}(A) = 0$ iff there exists a $r \times s$ zero submatrix of A with $r + s = n + 1$.

Proof. The proof goes by induction:

- Initialization:

  Let us consider a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $\mathrm{Per}(A)=0$. But $\mathrm{Per}(A)=ad + bc$ so

  $$(a, c) = (0, 0) \text{ or } (a, b) = (0, 0) \text{ or } (d, c) = (0, 0) \text{ or } (d, b) = (0, 0) \tag{46}$$

  In every case we have an 2×1 or 1×2 matrix full of zeros the the condition $r + s = 3 = n + 1$ is met

- Assume the lemma holds for all matrix size $k \times k$ with $k \leq n$. Let A be a $n + 1 \times n + 1$ matrix, such that $\mathrm{per}(A) = 0$. Two possible scenarios, as $\mathrm{per}(A) = \sum_{j=1}^{n} a_{i,j} \mathrm{per}(\Delta_{i,j})$ for all i, in the fist possible case we have that

  $$\exists i \forall j, a_{i,j} = 0 \tag{47}$$

  In that case we have a $1 \times n+1$ matrix of zeros so $r+s = 1+n+1 = n+2$.

  Otherwise for all i it exist a j such that $a_{i,j} \neq 0$. But the entries being all non-negative, all terms in sum of $\mathrm{per}(A)$ are non-negative we necessarily have that $\mathrm{per}(\Delta_{i,j}) = 0$. Since $\Delta_{i,j}$ is a $n \times n$ matrix we can apply it the

induction hypothesis. So there exists a $r \times s$ zero submatrix of $\Delta_{i,j}$ such that $r + s = n + 1$. By permuting rows and columns we can suppose that the zero matrix is in the right top corner of A such that:

$$A = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix} \tag{48}$$

Thank to the condition $r + s = k + 1$ the blocks B and D are square respectively $r \times r$ and $s \times s$. Moreover notice that

$$\text{per}(A) = \text{per}(B)\text{per}(D).$$

We either have $\text{per}(B) = 0$ or $\text{per}(D) = 0$. Let's treat the case were $\text{per}(B) = 0$ the other cases are handled similarly. If $\text{per}(B) = 0$ as B is an $r \times r$ matrix with $r \leq n$ we can use the induction hypothesis to deduce that there is an $i \times j$, matrix of zeros in B such that $i + j = r + 1$. By rearranging lings and columns in A, we finally have an $i \times (j + s)$ block of zeros with $i + j + s = r + 1 + s = k + 2$. This completes the proof.

□

Definition B.2. Doubly stochastic matrix [2, Page 504-506]
A *doubly stochastic matrix* is a non-negative square matrix $(x_{i,j})_{1 \leq i,j \leq n}$ that satisfies the following conditions:

$$\sum_{i=1}^{n} x_{i,j} = 1 \quad \text{and} \quad \sum_{j=1}^{n} x_{i,j} = 1.$$

Remark. The set of doubly stochastic matrices is exactly the set of kernels of reversible Markov chains.

Lemma B.2. [2, Page 504-506] Let $A \in \mathbb{R}^{n \times n}$ be a doubly stochastic matrix. Then $\text{per}(A) > 0$.

Proof. Let A be a bistochastic matrix of size $n \times n$ such that $\text{per}(A) = 0$. By the preceding lemma, there exists an $r \times s$ sub matrix of zeros such that $r + s = n + 1$. Without loss of generality we can assume that:

$$A = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix} \tag{49}$$

Let's denote $\Sigma(G)$ the sum of all entries in the matrix G. As A is bistochastic we have that

$$r = \Sigma(B) \leq \Sigma(B) + \Sigma(C) = n - s.$$

□

So $r + s \leq n$ which contradicts $r + s = n + 1$.

Definition B.3. Permutation matrix [2, Page 504-506]
A *permutation matrix* is matrix that has exactly one entry of 1 in each row and each column. All other entries are equal to zero.

Remark. There is a bijection between the set of all permutation matrices of size n and $\mathcal{S}_n$

$$\varphi : \begin{cases} \mathcal{S}_n \to \text{Perm} \\ \sigma \mapsto (\delta_{i,\sigma(i)})_{1 \leq i \leq n} \end{cases}$$

$$\varphi^{-1} : \begin{cases} \text{Perm} \to \mathcal{S}_n \\ P \mapsto (i \mapsto j \text{ si } p_{i,j} = 1) \end{cases}$$

.

We can then introduce the notation $P_\sigma = (\delta_{i,\sigma(i)})_{1 \leq i \leq n}$.

Theorem B.1. Birkhoff Von-Neumann [2, Page 504-506]
Every doubly stochastic matrix is a convex combination of permutation matrices.

Proof. If P is a permutation matrix, then the assertion is self-evident. If P is not a permutation matrix, then since P is a doubly stochastic matrix according to the Lemma B.2, we know that $\text{per}(P) > 0$. So, there necessarily exists a permutation $\sigma \in \mathcal{S}_n$ such that $1 > p_{1,\sigma(1)} \dots p_{n,\sigma(n)} > 0$. Let

$$\lambda_1 = \min\{p_{1,\sigma(1)}, \dots, p_{n,\sigma(n)}\}$$

and let Π_1 be the permutation matrix with 1's in the $i\sigma(i)$ position for all $1 \leq i \leq n$. Then

$$P_1 = \frac{P - \lambda_1 \Pi_1}{1 - \lambda_1}$$

is a doubly stochastic matrix with at least one more zero entry than P_1. We can decompose P as

$$P = \lambda_1 \Pi_1 + (1 - \lambda_1) P_1. \tag{50}$$

If P_1 is not a permutation we repeat the preceding steps ie there exists $0 < \lambda_2 < 1$ a permutation Π_2 such that,

$$P_2 = \frac{P_1 - \lambda_2 \Pi_2}{1 - \lambda_2}.$$

We repeat the process up until P_i is a permutation. This procedure as a finite amount of steps because we add at least a zero entry to P_i at each step and P_i has a finite number of entries. □

Remark. We can now see that the set of doubly stochastic matrices is exactly the convex hull of the permutation matrices. We will now refer to this set as the and denote it B_n.

B.2 Birkhoff's Polytope

In this subsection we will compute the dimension of B_n and the eigenvalues a of some Markov chain on the edges of B_3

Dimension

The dimension of B_n is the dimension of the embedding space minus the number of constraints. There are $2n$ constraints one being redundant. Indeed fixing the n row sums and the last $n-1$ column sums to 1,

$$1 \leq i \leq n, \sum_{j=1}^{n} x_{i,j} = 1, \quad 2 \leq j \leq n, \sum_{i=1}^{n} x_{i,j} = 1$$

then summing first on the rows then on the columns we get

$$\sum_{j=1}^{n} \sum_{i=1}^{n} x_{i,j} = n \tag{51}$$

but by inverting the order of summation we obtain

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_{i,j} = n - 1 + \sum_{j=1}^{n} x_{1,j} \tag{52}$$

So necessarily by combining (51) and (52) we get

$$\sum_{j=1}^{n} x_{1,j} = 1 \tag{53}$$

We could have done the same reasoning without loss of generality, on any other row or column. Proving that we only have $2n-1$ effective constraints so

$$\dim(B_n) = n^2 - 2n + 1.$$

Analysis of B_3

Here are the permutation matrices of B_3:

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, P_3 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$P_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, P_5 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, P_6 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Let us study the behavior of some Markov chains on the edges let's start with the edge $\{P_1, P_2\}$ Let $M \in \{P_1, P_2\}$, can be written as a convex combination

$$M = \lambda P_1 + (1-\lambda)P_2 = P_1 = \begin{pmatrix} \lambda & 1-\lambda & 0 \\ 0 & \lambda & 1-\lambda \\ 1-\lambda & 0 & \lambda \end{pmatrix}$$

Let us find his eigenvalues

$$\chi_M(X) = \left| \begin{pmatrix} X-\lambda & \lambda-1 & 0 \\ 0 & X-\lambda & \lambda-1 \\ \lambda-1 & 0 & X-\lambda \end{pmatrix} \right| = (X-\lambda)^3 - (\lambda-1)^3$$

So the eigenvalues are 1, $\lambda + (\lambda-1)e^{i\frac{2\pi}{3}}$,$\lambda + (\lambda-1)e^{-i\frac{2\pi}{3}}$. Thus any Markov chain with kernel M converges with speed $|\lambda + (\lambda-1)e^{-i\frac{2\pi}{3}}| = \sqrt{\lambda^2 - \lambda + 1}$.

The same holds for any convex combination of P_1 and P_3.

C Reminders

C.1 Point processes

The majority of what is written in this subsection comes from [6, Ch 2].

Heuristically, a point process is an at most countable random collection Z of points in some measured metric space S. We might think of Z as a mapping $\omega \mapsto Z(\omega)$ from Ω the space of all possible outcomes into the system of countable subsets of S where $(\Omega, \mathcal{F}, \mathbb{P})$ is an underlying probability space. Then Z can be identified with the family of mappings.

$$\omega \mapsto \eta(\omega, B) \coloneqq \#(Z(\omega) \cap B), \quad B \subset S.$$

We will now define precisely a point process.

If $\mathcal{S}$ is an $\sigma-$algebra on S we say the pair $(S, \mathcal{S})$ is a *measurable space*. In practice S will often be a Polish space[2] and $\mathcal{S} = \mathcal{B}(S)$ be the Borel $\sigma-$algebra on S. For the moment, we just assume that $(S, \mathcal{S})$ is a general measurable space.

Let us introduce the set of positive integer valued measures on S, $\mathbf{M}_{<\infty}(S)$,

$$\mathbf{M}_{<\infty}(S) \coloneqq \mathbf{M}_{<\infty} = \{\mu \text{ measure on } S \mid \forall B \in \mathcal{S}, \mu(B) \in \mathbb{N}\}.$$

and $\mathbf{M}(S)$ the set of countable sums of elements of $\mathbf{M}_{<\infty}$.

$$\mathbf{M}(S) \coloneqq \mathbf{M} = \left\{ \sum_{n=1}^{\infty} \mu_n \mid \forall n \in \mathbb{N}, \mu_n \in \mathbf{M}_{<\infty} \right\}.$$

As an example of an element of $\mathbf{M}$, we have the zero measure 0, which is identically zero on $\mathcal{S}$. Another example is the *Dirac measure* δ_x at a point $x \in S$ given by $\delta_x(B) \coloneqq \mathbf{1}_B(x)$.

More generally, any sequence (finite or not) $(x_n)_{n \geq 1}$ can be used to define a measure

$$\mu = \sum_{k=1}^{\infty} \delta_{x_n}, \tag{54}$$

then $\mu \in \mathbf{M}$ and

$$\mu(B) = \sum_{n=1}^{\infty} \mathbf{1}_B(x_n), B \in \mathcal{S}. \tag{55}$$

And more generally, for any measurable $f : S \to \mathbb{R}_+$, we have

$$\int f d\mu = \sum_{n=1}^{\infty} f(x_n). \tag{56}$$

[2] In a Polish space the study of Borel measures is convenient, Prokhorov's theorem insures convergence of sequences of measures, and separability ensures that the Borel sigma algebra is large enough to distinguish every pair of points. In our specific case in a polish space any Poisson process is proper.

In general, it cannot be guaranteed that every $\mu \in \mathbf{M}$ can be written in the form (54). However, it is sufficient for μ to be σ-finite for this to be true.

Definition C.1. [6, Ch.2] ν is an *s-finite* measure on S if ν is a countable sum of finite measure on S.

Definition C.2. [6, Ch.2] ν is an *σ-finite* measure on S there is a sequence $B_m \in \mathcal{S}, m \in \mathbb{N}$ such that

$$\bigcup_m B_m = S \quad \text{and} \quad \forall m \in \mathbb{N}, \nu(B_m) < \infty.$$

Let $\mathcal{M}(S) \equiv \mathcal{M}$ be the σ-algebra generated by the sets

$$\{\mu \in \mathbf{M} : \mu(B) = k\}, B \in \mathcal{S}, k \in \mathbb{N}$$

i.e., $\mathcal{M}$ is the smallest σ-algebra making $\mu \mapsto \mu(B)$ measurable for all $B \in \mathcal{S}$.

Definition C.3. Point process [6, Ch.2]
A *point process* on S is a random element η of $(M, \mathcal{M})$ which is a mapping $\eta : \Omega \to \mathbf{M}$.

Example. Let $\mathbb{Q}$ be a probability measure on S and suppose that $X_1, \ldots, X_n$ are independent random elements in S with distribution $\mathbb{Q}$.

Then

$$\eta := \delta_{X_1} + \cdots + \delta_{X_n}$$

is a point process on S. Because

$$\mathbb{P}(\eta(B) = k) = \binom{n}{k} \mathbb{Q}(B)^k (1 - \mathbb{Q}(B))^{n-k}, \quad k = 0, \ldots, n,$$

η is referred to as a *binomial process* with sample size n (see example 2.3 in [6, Page 11]) and disorder $\mathbb{Q}$ (see page 10 of [3] for motivation of the term disorder) [3, Page 10].

Definition C.4. Proper point process [6, Ch.2]
A point process η on S is a *proper point process* if there exists random elements $X_1, X_2, \ldots \in S$ and an $\bar{\mathbb{N}} := \mathbb{N} \cup \infty$ valued random variable κ such that almost surely

$$\eta = \sum_{n=1}^{\kappa} \delta_{X_n}.$$

Remark. The class of proper point process is very large. Indeed if S is a Borel subspace of a complete separable metric space, then any locally finite point process ($\mathbb{P}(\eta(B) < \infty) = 1$ for every bounded $B \in \mathcal{S}$) on S is proper.

Definition C.5. Intensity measure [6]
The *intensity measure* of a point process η on S is the measure λ defined by

$$\lambda(B) := \mathbb{E}[\eta(B)], \quad B \in \mathcal{S}.$$

C.2 Poisson processes

Definition C.6. Poisson process [6]
Let λ be an s-finite measure on S. A *Poisson process* with intensity measure λ is a point process η on S with the following two properties:

- For every $B \in \mathcal{S}$ the distribution of $\eta(B)$ is Poisson with parameter $\lambda(B)$;
- For every $m \in \mathbb{N}$ and all pair wise disjoint sets $B_1, \ldots, B_n \in \mathcal{S}$ the random variables $\eta(B_1), \ldots, \eta(B_n)$ are independent.

Notation. If S is a is a Borel subspace of a complete separable metric space, then any Poisson point process on that space is proper. In that case we will often denote by $\Pi = \{X_n, n \leq \eta(S)\}$ the ensemble of random elements defined in C.4. We also have that $\forall B \in \mathcal{S}, \eta(B) = \#\{\Pi \cap B\}$.

Proposition C.1. [6] Let η be a Poisson process on S with intensity measure λ satisfying $0 < \lambda(S) < \infty$. Then the conditional distribution $\mathbb{P}(\eta \in \cdot \mid \eta(S) = n), n \in \mathbb{N}$ is that of a binomial process with sample size n and sampling distribution $\mathbb{Q} := \frac{\lambda}{\lambda(S)}$.

Proof. Let $n \in \mathbb{N}$,$A \in \mathcal{M}$, we denote $\mathbb{P}_n(A) = \mathbb{P}(A \mid \eta(S) = n)$. It is sufficient to take A of the form $A = \{\mu \in \mathbf{M} : \mu(B) = k\}, B \in \mathcal{S}, k \leq n$. Let's compute $\mathbb{P}_n(\eta \in A)$.

Using Bayes theorem:

$$\mathbb{P}_n(\eta \in A) = \frac{\mathbb{P}(\{\eta \in \{\mu \in \mathbf{M} : \mu(B) = k\}\} \cap \{\eta(S) = n\})}{\mathbb{P}(\eta(S) = n)}. \tag{57}$$

We then show that the event $\{\eta \in \{\mu \in \mathbf{M} : \mu(B) = k\}\} \cap \{\eta(S) = n\}$ is equal to $\{\eta(B) = k\} \cap \{\eta(B^c) = n - k\}$. Let us first notice that

$$\{\eta \in \{\mu \in \mathbf{M} : \mu(B) = k\}\} = \{\eta(B) = k\}. \tag{58}$$

Since $S = B \cup B^c$ we can rewrite $\{\eta(S) = n\}$ as follows:

$$\{\eta(S) = n\} = \{\eta(B \cup B^c) = n\}.$$

as η is a measure and $B \cap B^c = \emptyset$ we get

$$\{\eta(S) = n\} = \{\eta(B) + \eta(B^c) = n\}. \tag{59}$$

Then, combining (58) and (59) we get

$$\{\eta \in \{\mu \in \mathbf{M} : \mu(B) = k\}\} \cap \{\eta(S) = n\} = \{\eta(B) = k\} \cap \{\eta(B) + \eta(B^c) = n\}.$$

We can further simplify as follows

$$\{\eta \in \{\mu \in \mathbf{M} : \mu(B) = k\}\} \cap \{\eta(S) = n\} = \{\eta(B) = k\} \cap \{\eta(B^c) = n-k\}. \quad (60)$$

Putting (57) and (60) together we get

$$\mathbb{P}_n(\eta \in A) = \frac{\mathbb{P}(\{\eta(B) = k\} \cap \{\eta(B^c) = n-k\})}{\mathbb{P}(\eta(S) = n)}.$$

Since $B \cap B^c = \emptyset$ we have that $\eta(B)$ and $\eta(B^c)$ are independent. Thus

$$\mathbb{P}_n(\eta \in A) = \frac{\mathbb{P}(\{\eta(B) = k\})\mathbb{P}(\{\eta(B^c) = n-k\})}{\mathbb{P}(\eta(S) = n)}.$$

But $\forall A \in \mathcal{S}, \eta(A) \sim \mathcal{P}(\lambda(A))$, so

$$\mathbb{P}_n(\eta \in A) = \frac{\frac{(\lambda(B))^k e^{-\lambda(B)}}{k!} \frac{(\lambda(S)-\lambda(B))^{n-k} e^{-(\lambda(S)-\lambda(B))}}{(n-k)!}}{\frac{(\lambda(S))^n e^{-(\lambda(S))}}{n!}}.$$

With tedious but elementary simplifications we obtain

$$\mathbb{P}_n(\eta \in A) = \binom{n}{k} \frac{(\lambda(B))^k (\lambda(S) - \lambda(B))^{n-k}}{\lambda(S)^n}.$$

We finally get

$$\mathbb{P}_n(\eta \in A) = \binom{n}{k} (\mathbb{Q}(B))^k (1 - \mathbb{Q}(B))^{n-k}.$$

□

Proposition C.2. Joint distribution under $\mathbb{P}_n$
Under the same assumptions as C.1 let $A_1, \ldots, A_k$ be in $\mathcal{S}$, and $i_0, i_1, \ldots, i_k$ be positive integers such that $\sum_{p=1}^{k} i_p = n$ then

$$\mathbb{P}_n(\eta(A_1) = i_1, \ldots, \eta(A_k) = i_k) = \frac{n!}{i_0! i_1! \ldots i_n!} \prod_{p=1}^{k} \mathbb{Q}(A_p)^{i_p}$$

Theorem C.1. Mecke equation [6]
Let λ be an $s-$finite measure on S and η a point process on S. Then η is a Poisson process with intensity measure λ if and only if

$$\mathbb{E}\left[\int f(x, \eta)\eta(dx)\right] = \int \mathbb{E}\left[f(x, \eta + \delta_x)\right] \lambda(dx), \quad (61)$$

for all $f : S \times \mathbf{M} \longrightarrow \bar{\mathbb{R}}_+$.

Theorem C.2. Mapping Theorem [6]
Let η be a point process on S with intensity measure λ and let $T : S \to S'$ were $(S', \mathcal{S}')$ is a measurable space. Then $T(\eta)$ is a point process with intensity measure $\lambda \circ T^{-1}$. If η is a Poisson process, then $T(\eta)$ is a Poisson process too.

Example. Inverse CDF
Let X be a real valued random variable a of CDF F, and U a uniform random variable on $[0, 1]$. We then know that $F^{-1}(U) \sim X$ where F^{-1} is the generalized inverted CDF:

$$F^{-1}(y) = \inf\{x \in \mathbb{R} \mid F(x) \geq y\}.$$

By the mapping theorem, we know that η is an homogeneous Poisson process on $[0, 1]$. If μ is a probability measure on $\mathbb{R}$, we can define $F(x) = \mu(]-\infty, x])$ then by the mapping theorem $F^{-1}(\eta)$ is a Poisson process of intensity measure μ. So we can simulate any Poisson process on $\mathbb{R}$ from an homogeneous Poisson process on $[0, 1]$.

References

[1] Rainer E. Burkard, Mauro Dell'Amico, and Silvano Martello. *Assignment problems*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 2009.

[2] Harry Dym. *Linear algebra in action*. Number volume 78 in Graduate studies in mathematics. American mathematical society, Providence (R.I.), 2007.

[3] Matteo d'Achille. *Statistical Properties of the Euclidean Random Assignment Problem*. Phd thesis, Université Paris-Saclay, 2020.

[4] Valentin Féray. Processus stochastiques discrets – méthode de moments et marche aléatoire. Lecture notes (in French).

[5] Harold W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

[6] Günter Last and Mathew Penrose. *Lectures on the Poisson Process*. Cambridge University Press, 2017.

[7] Chandra Nair, Balaji Prabhakar, and Mayank Sharma. Proofs of the Parisi and Coppersmith-Sorkin conjectures for the finite random assignment problem. *IEEE*, pages 168–178, 2003.

[8] Randall Pyke. Spacings. *Journal of the Royal Statistical Society Series B*, 27(3):395–436, 1965.

[9] Michel Talagrand. Matching random samples in many dimensions. *The Annals of Applied Probability*, 2(4):846–856, 1992.