

Master 2 research thesis proposal

RANDOM ASSIGNMENT PROBLEMS FOR SATELLITE DATA ACQUISITION

Matteo D'ACHILLE & Jessie LEVILLAIN

matteo.d-achille@univ-lorraine.fr jessie.levillain@cnes.fr

December 10, 2025

Introduction: Assignment between two sets of points is a recurring problem in satellite image processing. For example, matching corresponding points in two different satellite images is needed for georeferencing one image using . for example, georeference an image using another image as a reference or for 3D reconstruction through stereo-photogrammetry. Monitoring surface water dynamics is a crucial issue in the field of Earth Observation (EO), as it has many applications in various contexts, ranging from renewable energy production to climate change, or anthropic-related concerns. Data from EO satellites (such as Sentinel-1, Sentinel-2, Sentinel-3, and SWOT) is currently used to estimate surfaces area and depth of lakes and reservoirs in real-time. However, each data source has its limitations, such as its specific uncertainties depending on measuring techniques or sampling. Image processing techniques, such as Markov Random Field [6] or its extension, Conditional Random Field [7], have been used over satellite data to extract either 2D or 3D contour lines of a water body [2, 5]. The aim of this project is to determine the minimal admissible sampling frequency by looking at those contour lines, and its influence on the resulting data.

Given a grid, we wish to track the dynamics of many identical agents, materializing the contour \mathcal{C} of a water body. n points on \mathcal{C} are modeled at time t by a family of blue points, $\mathcal{B} = \{b_i\}_{i=1}^n$. The same n points on \mathcal{C} at another fixed time t' are modeled by red points $\mathcal{R} = \{r_i\}_{i=1}^n$. \mathcal{B} and \mathcal{R} are independent $\text{Binom}(\nu, n)$ point processes, where ν is a given intensity measure on an euclidean metric space $(\mathcal{M}, \mathcal{D})$. The assignment of b_i to r_j is mapped by the following cost function

$$c : \begin{cases} \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R} \\ (b_i, r_j) \mapsto c_{ij} = c(b_i, r_j). \end{cases}$$

A configuration of the whole system can then be encoded by a permutation $\pi \in \mathcal{S}_n$, with the associated permutation matrix P_π . The energy of that configuration is then denoted by

$$\mathcal{H}(\pi) = \sum_{i=1}^n c_{i\pi(i)} = \text{Tr}(P_\pi c).$$

An optimal assignment π_{opt} minimizes the energy $\mathcal{H}_{opt} := \mathcal{H}(\pi_{opt}) = \min_{\pi \in \mathcal{S}_n} \mathcal{H}(\pi)$, which is a random variable. The ground state energy \mathcal{H}_{opt} is proportional to the p -Wasserstein distance associated to \mathcal{B} and \mathcal{R} . This minimization problem is called an Euclidean Random Optimization Problem (ERAP).

Problems:

- *Theoretical aspects on first and second moments in 2d:* It is established [1] that $\mathbb{E}[\mathcal{H}_{opt}] \sim \frac{1}{2\pi} \log n$ in dimension $d = 2$. We would like to investigate further this expansion, for example by comparing two domains Ω and Ω' , via $\mathbb{E}[\mathcal{H}_{opt}^\Omega - \mathcal{H}_{opt}^{\Omega'}]$. Another possibility would be to study the variance, $\text{Var}[\mathcal{H}_{opt}]$.
- *Discrete Fourier analysis in 2D:* Considering now that \mathcal{B} are nodes of a 2D grid, \mathcal{R} still being a binomial process, we would like to use Fourier analysis to study the different contributions to $\mathbb{E}[\mathcal{H}_{opt}]$. This can then be linked directly to the last point.
- *Application to satellite data acquisition:* The Hungarian method solves the ERAP in time complexity $\mathcal{O}(n^3)$ in the worst case [9]. The goal is to investigate this algorithm, along with other ones polynomial in complexity (primal-dual, Jonker-Volgenant...) and apply them to the ERAP on satellite data. The previous theoretical results will also be tested numerically on CNES data.

Keywords: ERAPs, Monge-Kantorovich problem, Hungarian method, discrete Fourier analysis.

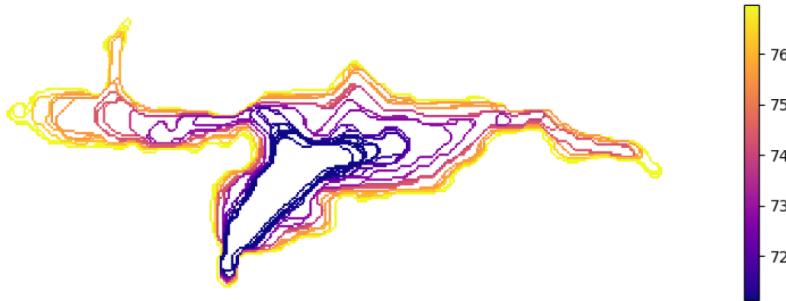


Figure 1: Extracted contours measuring bathymetry (in m) on Vioreau lake (44) from 3D segmentation on SWOT data [2].

Candidate profile: Master student and/or 3rd-year student in an engineering school, with a strong background in applied mathematics, in particular in probability theory and optimal transport. Coding skills in Python are expected from the candidate, no prior knowledge in hydrology or EO is required.

What we offer: The project lasts in between 4 and 6 months and is hosted at the Laboratoire de Mathématiques d'Orsay with a flexible starting date in April 2025. Possible visits at the CNES headquarters in Toulouse.

Application: We require a CV and cover letter explaining your interest in the project, via email to both supervisors.

References

- [1] L. Ambrosio, M. Goldman, D. Trevisan, On the quadratic random matching problem in two-dimensional domains, *Electronic Journal of Probability*, Electron. J. Probab. 27(none), 1-35, 2022.
- [2] R. Cargnello, *Hybridization of multi-sensor and multi-temporal data for monitoring water resources*, internship report, CNES, 2024.
- [3] M. Chertkov, L. Kroc, F. Krzakala, M. Vergassola and L. Zdeborová, *Inference in particle tracking experiments by passing messages between images*, Proceedings of the National Academy of Sciences, 2010.
- [4] M. D'Achille, *Statistical Properties of the Euclidean Random Assignment Problem*, PhD thesis, Paris-Saclay University, 2020. <https://hal.archives-ouvertes.fr/tel-03098672>
- [5] N. Gasnier, *Use of multi-temporal and multi-sensor data for continental water body extraction in the context of the SWOT mission*, PhD thesis, Institut polytechnique de Paris, 2022.
- [6] S. Geman and D. Geman, *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pp. 721-741, 1984.
- [7] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, In: Proceedings of the Eighteenth International Conference on Machine Learning. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 282–289, 2001.
- [8] M. Moharrami, C. Moore, J. Xu, *The planted matching problem: Phase transitions and exact results*, The Annals of Applied Probability, Ann. Appl. Probab. 31(6), 2663-2720, 2021.
- [9] J. Munkres, Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics* 5, 32–38, 1957.
- [10] T. Quarck, Random assignment problems for satellite data acquisition, internship report, Sorbonne Université, 2025

0.1 Réunion du 09/01/2025

On étudie le problème d'assignation en dimension $d = 2$, sur Ω une variété Riemanienne compacte. On prend comme fonction de coût la distance euclidienne, donc $\mathcal{H}_{opt} = \min_{\pi \in S_n} \sum_{i=1}^n |b_i - r_{\pi(i)}|^2$. On sait qu'on peut écrire

$$\mathcal{H}_{opt} =_{n \rightarrow \infty} \frac{1}{2\pi} \log n + R_\Omega(n), \quad (1)$$

tel que $R_\Omega(n) \leq \sqrt{\log \log n}$.

On a ensuite plusieurs pistes de travail dans le cadre du stage (et même qui dépassent ce cadre).

0.1.1 Estimation de $R_\Omega(n)$ (en dehors du stage)

Actuellement, une conjecture stipule que $R_\Omega(n) = c_\Omega + O(1)$, avec c_Ω une constante qui dépend que du domaine, et qui peut être calculée comme $c_\Omega = \text{Res}(\zeta_\Omega^{(s)}, s = 1)$, où

$$\zeta_\Omega^{(s)} = \sum_{\lambda \in Sp(\Delta_\Omega)} \frac{1}{\lambda^s}. \quad (2)$$

L'ensemble $Sp(\Delta_\Omega)$ est celui des valeurs propres de Δ_Ω avec conditions aux bords de Neumann.

0.1.2 Comparaison de deux domaines Ω et Ω'

On prend par exemple Ω' avec les mêmes réalisations de \mathcal{B} et \mathcal{R} que dans Ω , mais on impose des conditions périodiques aux bords gauche et droit (cela change la notion de distance, cf Figure 2).

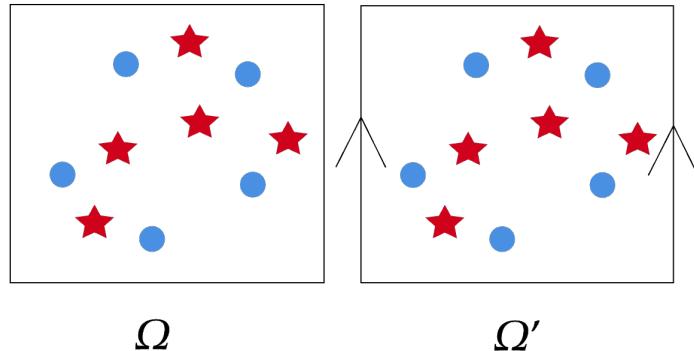


Figure 2: Les deux domaines choisis.

On s'intéresse à l'espérance de la différence des deux énergies des états fondamentaux, et on veut montrer que

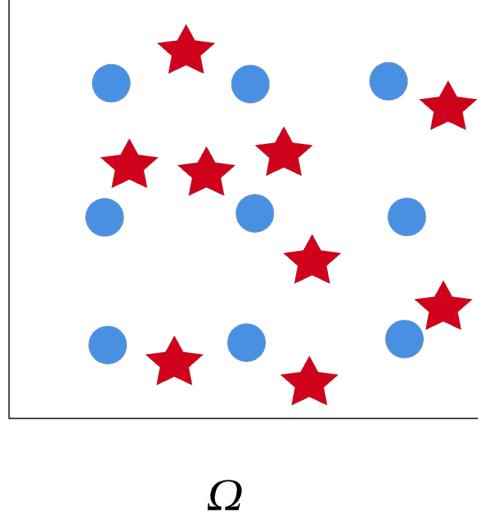
$$\mathbb{E}[\mathcal{H}_{opt}^\Omega - \mathcal{H}_{opt}^{\Omega'}] \rightarrow_{n \rightarrow \infty} c_\Omega - c_{\Omega'}. \quad (3)$$

0.1.3 Estimation de la variance

Montrer que $\text{Var}[\mathcal{H}_{opt}] = c_\Omega \simeq \sum_{(n,m) \in (\mathbb{Z}*)^2} \frac{1}{n^2 + cm^2}$.

0.1.4 Analyse de Fourier discrète

Les points \mathcal{B} sont maintenant des nœuds d'une grille en 2D, les points \mathcal{R} sont toujours un processus binomial, comme illustré en Figure 3.



Ω

Figure 3: Grille aux nœuds bleus.

On définit, pour $i \in \{1, \dots, n\}$, $\mu_i = r_{\pi(i)} - b_i$, et donc on peut écrire l'énergie de l'état fondamental comme $\mathcal{H}_{opt} = \|\mu\|^2$. Grâce à la décomposition de Helmholtz-Hodge, on peut trouver φ et ψ tels que

$$\mu = \nabla_{dis}\varphi + \nabla_{dis} \times \psi. \quad (4)$$

On a alors deux conjecture qu'on aimeraient prouver:

1. φ et ψ sont disjoints.
2. Quand $n \rightarrow \infty$, $\mathbb{E}[\|\nabla\varphi\|^2] \sim \frac{1}{2\pi} \log n$ et $\mathbb{E}[\|\nabla \times \psi\|^2] \rightarrow c_\Omega$.

[Matteo: Merci beaucoup du résumé! Vendredi non plus j'ai reçu une réponse definitive sur la question administrative. Mais le lundi est obligatoire, car tout le monde sera là en raison de la visite de l'HCERES:]

[Matteo: PS: voici le lien à la repository github <https://github.com/matteodachille/ERAPs2d>] [Jessie: Merci pour le repository! Super, j'espère que tu réussiras à avoir une réponse aujourd'hui. Si c'est pas le cas, on fait quoi, on commence par l'envoyer que dans les ENS (pour pas être trop en retard)?] [Matteo: Me re-voici ! Avec la visita de l'HCERES tout le monde est très occupé ... mais j'ai des éléments de plus: first of all, le fond du problème est le suivant : quoi qui ce soit (stage ou mémoire), il faut un permanent comme responsable. Or, je ne crois pas qu'il y aura de problème à demander à Nicolas, par contre, il y a une question thématique, car le thème n'est pas trop prêt de ses thèmes de recherche. Du coup il vaudrait mieux éviter que l'étudiant.e retenu.e puisse imaginer une thèse avec lui derrière...; bref, on est dans une petite impasse. Voici ce que je propose: pour l'instant on ne publie pas de sujet (en tout cas j'aurais pas le droit sans validation officielle du

directeur des études, car ça risque d'invalider le mémoire/stage...), mais on va essayer de recruter un.e normalien.ne/polytechnicien.ne "dans les couloirs". Avec un peu de chance, plusieurs étudiant.e.s qui entendent la rumeur nous contacteront et avec ça en main on essayera de sortir de cette impasse. Qu'en dis-tu ? Si t'es d'accord, je peux lancer de suite mes contacts à Ulm.] [Jessie: Donc si je résume le plan: en recrutant quelqu'un dans les couloirs, la personne serait au courant qu'il n'y aura pas de thèse sur le même sujet avec Nicolas donc ce serait ok ? Parallèlement, tu demandes quand même à Nicolas d'être prête-nom aussi pour que le mémoire soit valide ? Ca me va très bien si tu fais passer le mot à des étudiants, voire que les profs parlent d'un sujet avec nous et que sur le papier c'est Nicolas le "chef". Du coup, quand tu as un étudiant intéressé tu lui envoies le PDF, sans le mettre sur aucun site perso pour l'instant c'est bien ça ? Je peux pas vraiment le diffuser comme stage à l'ENS Paris-Saclay via mes directeurs de thèse, vu que ça arriverait sur une liste "officielle" de stages... Du coup je me repose sur toi, sauf si ça évolue à un moment et qu'on peut le publier !][Matteo: Oui, il faut leur dire qu'il y a pas de financement de thèse derrière ! Concernant Nicolas, oui. Je pense que c'est mieux de garder le sujet que pour nous pour l'instant. De toute façon on a un peu d'avance sur le timing!] [Jessie: Super, tiens moi au courant de ce que ça donne à Ulm et/ou dans les couloirs d'Orsay alors. Je te fais confiance niveau timing, parce que je sais que nous à Cachan c'était pas du tout le même rythme (ils voulaient qu'on ait trouvé avant fin janvier quand j'étais à l'ENS). Hâte de voir qui mordra à l'hameçon !] [Matteo: yes ;)]