

# Influence de l'Echantillonnage sur la Distance de Wasserstein Empirique

December 2, 2022

## Superviseurs

- Laetitia Chapel, Professeur Assistant, équipe Obelix team.  
Principal thème de recherche : Transport Optimal pour le Machine Learning
- Guillaume Mahey, Doctorant, équipe Obelix .  
Principal thème de recherche : Transport Optimal pour le Machine Learning

**Mots Clés** Optimal transport, Distance de Wasserstein, Échantillonnage

## Contexte

Le Transport Optimal [Vil09] [PC<sup>+</sup>19] fût introduit en 1781 par Monge [Mon81], puis développé durant la seconde guerre mondiale par Kantorovich [Kan42]. C'est un domaine de recherche de mathématiques moderne, mis en lumière par des acteurs comme Villani et Figalli, médailles Fields 2010 et 2018.

En Transport Optimal, la quantité fondamentale est la distance de Wasserstein. Elle est définie comme la façon la plus économe de transporter de la masse d'une distribution vers une autre. Cette quantité définit une distance dans l'espace des probabilités.

Soit  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  et  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  un coût de transport

$$W(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\pi(x, y) \quad (1)$$

avec  $\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \text{ tel que } \pi(\mathbb{R}^d \times A) = \nu(A) \text{ et } \pi(A \times \mathbb{R}^d) = \mu(A) \forall A \text{ mesurable de } \mathbb{R}^d\}$

Le transport optimal s'est récemment imposé dans divers domaines d'applications (Machine Learning, Finance, Traitement d'images, Biologie...). Dans ces domaines, les objets d'intérêt (nuages de points, images, graphes,...) peuvent être encodés par des distributions. En ce sens, la distance de Wasserstein peut être utilisée pour comparer ces objets.

Dans les applications du transport optimal, les distributions sont souvent estimées empiriquement. La distance de Wasserstein est alors approximée par son équivalent empirique.

Soit  $\mu_n$  (resp.  $\nu_n$ ) estimation empirique de  $\mu$  (resp.  $\nu$ ), avec  $\mu_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mu$  (resp.  $\nu_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \nu$ ), l'estimation empirique de  $W(\mu, \nu)$  est donnée par  $\widehat{W}(\mu_n, \nu_n)$  avec :

$$\widehat{W}(\mu_n, \nu_n) \xrightarrow[n \rightarrow \infty]{} W(\mu, \nu) \quad (2)$$

La manière d'échantillonner (i.e la manière d'estimer empiriquement  $\mu$  et  $\nu$ ) est un levier sur les propriétés de convergence de  $\widehat{W}$ . Par exemple, la vitesse de convergence [WB19] et la variance [RMS16] de  $\widehat{W}$  sont des propriétés qui diffèrent selon la façon d'échantillonner. **L'objectif de ce stage est de**

quantifier l'évolution des propriétés de l'estimateur empirique de Wasserstein en fonction des différentes méthodes d'échantillonnage.

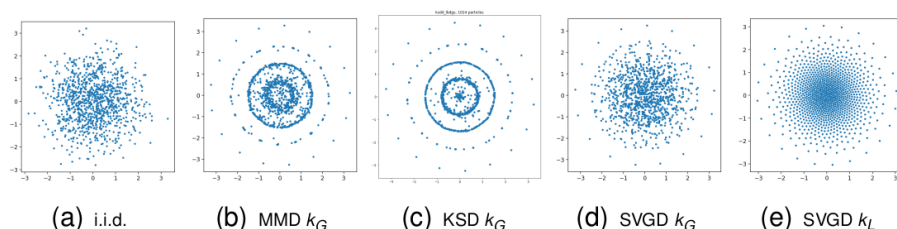


Figure 1: Différents Echantillonnage d'une Gaussienne [XKS22]

Classiquement, les échantillonnages de loi Gaussienne  $X \sim \mathcal{N}$ , sont réalisés avec la méthode Box-Muller ou la méthode Ziggurat [Box58] [MT00]. Ces 2 méthodes transforment des échantillonnages uniformes vers des Gaussiennes. La première en utilisant la fonction de répartition  $F_X$  et la seconde avec la densité  $f_X$ . Ces deux méthodes sont déjà implémentées en Python, avec le package numpy.

Plus récemment des méthodes de système de particules émergent comme méthodes d'échantillonnage [AKSG19] [KAFMA21]. Partant d'une fonction de dissimilarités entre distributions, on peut construire une fonction de perte  $\mathcal{F}$  tel que le problème d'échantillonnage revient à un problème d'optimisation avec  $X \sim \arg \min_{\mu \in \mathcal{P}_2} \mathcal{F}(\mu)$ . Ce problème de minimisation se résout à l'aide des flows de gradient Wasserstein [San17] ou avec la descente de gradient de Stein [LW16]. Ces méthodes de résolution sont les équivalents de la descente de gradient dans l'espace des probabilités. Toutes ces méthodes ont déjà été implémentées et sont disponibles sur le github des auteurs.

Mieux comprendre l'impact de l'échantillonnage sur la distance de Wasserstein empirique permettrait d'évaluer les bornes théoriques de cet estimateur. Cela serait également un premier pas vers la construction d'un traitement des données préalable à l'estimation empirique. Ce traitement serait construit afin de tirer de meilleures propriétés sur  $\widehat{W}$  (accélérer la convergence par exemple).

## Objectifs du Stage

Le stage se déroulera en 3 étapes, l'étudiant-e devra :

1. Se familiariser avec le transport optimal et son implémentation numérique
2. Se familiariser avec les méthodes d'échantillonnage, aucune implémentation n'est demandée
3. Quantifier les propriétés de  $\widehat{W}$  selon la méthode d'échantillonnage

## Compétences requises et profil de l'étudiant-e

L'étudiant-e devra avoir un bagage mathématique suffisant en analyse, probabilités et théorie de la mesure. L'étudiant-e devra également être familier.e avec le langage de programmation Python, il-elle sera amené-e à utiliser la librairie POT <https://pythonot.github.io/> [FCG<sup>+</sup>21].

## References

- [AKSG19] Michael Arbel, Anna Korba, Adil Salim, and Arthur Gretton. Maximum mean discrepancy gradient flow. *Advances in Neural Information Processing Systems*, 32, 2019. <https://github.com/MichaelArbel/MMD-gradient-flow>.
- [Box58] George EP Box. A note on the generation of random normal deviates. *Ann. Math. Statist.*, 29:610–611, 1958.

- [FCG<sup>+</sup>21] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boissunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *J. Mach. Learn. Res.*, 22(78):1–8, 2021.
- [KAFMA21] Anna Korba, Pierre-Cyril Aubin-Frankowski, Szymon Majewski, and Pierre Ablin. Kernel stein discrepancy descent. In *International Conference on Machine Learning*, pages 5719–5730. PMLR, 2021. <https://github.com/pierreablin/ksddescent>.
- [Kan42] Leonid V Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.
- [LW16] Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in neural information processing systems*, 29, 2016.
- [Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.
- [MT00] George Marsaglia and Wai Wan Tsang. The ziggurat method for generating random variables. *Journal of statistical software*, 5:1–7, 2000.
- [PC<sup>+</sup>19] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
- [RMS16] Thomas Rippl, Axel Munk, and Anja Sturm. Limit laws of the empirical wasserstein distance: Gaussian distributions. *Journal of Multivariate Analysis*, 151:90–109, 2016.
- [San17] Filippo Santambrogio. {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences*, 7(1):87–154, 2017.
- [Vil09] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [WB19] Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.
- [XKS22] Lantian Xu, Anna Korba, and Dejan Slepcev. Accurate quantization of measures via interacting particle-based optimization. In *International Conference on Machine Learning*, pages 24576–24595. PMLR, 2022. <https://github.com/xulant/accurate-quantization-and-nsvgd>.