

Approches variationnelles pour le stippling : distances L^2 ou transport optimal ?

Frédéric DE GOURNAY^{1,2}, Jonas KAHN^{1,2}, Léo LEBRAT¹, Pierre WEISS^{1,2}

¹Institut de Mathématiques de Toulouse
CNRS UMR 5219, Université de Toulouse, France.

²Équipe PRIMO, ITAV
CNRS USR 3505, Université de Toulouse, France.

frederic.de-gournay@insa-toulouse.fr , jonas.kahn@math.univ-lille1.fr
lebrat@insa-toulouse.fr, pierre.armand.weiss@gmail.com

Résumé – Le stippling est un problème qui a beaucoup progressé dernièrement grâce à l’introduction de méthodes variationnelles. On s’intéresse ici à deux types de formulations. L’une repose sur une distance L^2 entre mesures et fait appel à des outils d’analyse harmonique appliquée. L’autre repose sur la distance de Wasserstein et fait appel à des outils de géométrie algorithmique. Différentes méthodes de résolution et de discrétisation sont comparées et nous présentons leurs atouts et leurs limitations.

Abstract – Stippling is a problem that recently found elegant and efficient solutions thanks to the introduction of variational methods. The aim of this paper is to compare two state-of-the-art approaches: one is based on the minimization of an L^2 norm (with links to applied harmonic analysis), while the other is based on the Wasserstein distance (with links to computational geometry).

1 Introduction

Le stippling consiste à approcher une mesure de probabilité cible μ sur $\Omega \subset \mathbb{R}^2$ par une mesure discrète de la forme $\nu = \nu(\mathbf{p}) = \frac{1}{n} \sum_{i=1}^n \delta_{p_i}$, avec des positions $\mathbf{p} = (p_i)_{1 \leq i \leq n} \in \Omega^n$. Le principe des méthodes variationnelles est de trouver un ensemble de positions \mathbf{p} qui minimise une certaine distance $d(\mu, \nu)$ entre μ et ν :

$$\min_{\mathbf{p}} d(\mu, \nu(\mathbf{p})). \quad (1)$$

Deux types de distance ont été proposées récemment.

Distances L^2 . La première [9, 3] repose sur une distance L^2 , après régularisation par un noyau de convolution. Elle est définie par :

$$d_{L^2}(\mu, \nu(\mathbf{p})) = \frac{1}{2} \|h \star (\mu - \nu(\mathbf{p}))\|_2^2, \quad (2)$$

où $h \in C^0(\Omega)$ est un noyau régularisant.

Distances de transport. La seconde [5] utilise la distance W_2 de Wasserstein (aussi appelée distance de transport) définie par :

$$d_{W_2}(\mu, \nu) = \left(\inf_{\gamma \in \Pi(\nu, \mu)} \int_{\Omega^2} \|x - y\|^2 d\gamma(x, y) \right)^{\frac{1}{2}}, \quad (3)$$

où Π est l’ensemble des couplages entre μ et ν , i.e. l’ensemble des mesures sur Ω^2 dont les marginales sur x et y sont respectivement μ et ν .

Contribution. Cet article est une étude de l’efficacité de ces deux distances d’un point de vue numérique et qualitatif. Nous décrivons plusieurs algorithmes, dont certains sont originaux. Ils font appel à des outils très différents tels que l’analyse harmonique appliquée (transformées de Fourier non uniformes), la géométrie algorithmique (diagrammes de Laguerre), l’optimisation convexe et non convexe (méthodes proximales). Ces outils sont utilisés par des communautés différentes qui s’ignorent partiellement. Notre ambition est d’aider à rapprocher ces communautés en popularisant certaines approches encore méconnues par un grand nombre de chercheurs en traitement du signal.

2 Méthodes numériques

Sous une apparente simplicité, la classe de problèmes (1) est en fait extrêmement complexe. Par exemple, elle contient le problème de Thomson (placer des points de façon “uniforme” sur une sphère) qui est dans la liste de Smale [10] des problèmes ouverts de mathématiques à résoudre pour le XXIème siècle. Nous détaillons ci-dessous les méthodes numériques émergentes pour le résoudre lorsque le nombre n de diracs est grand. Dans le cas de la distance de transport, nous présenterons trois approches : discrète, discrète régularisée et semi-discrète.

2.1 Distance de convolution

Théorie Lorsque $d = d_{L^2}$, le problème (1) peut se réécrire sous la forme d'un problème non convexe sur les positions \mathbf{p} :

$$\min_{\mathbf{p} \in \Omega^n} \underbrace{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n H(p_i - p_j)}_{F(\mathbf{p}) \equiv \text{Répulsion}} - \underbrace{\frac{1}{n} \sum_{i=1}^n \int_{\Omega} H(x - p_i) d\mu(x)}_{G(\mathbf{p}) \equiv \text{Attraction}}, \quad (4)$$

où H est défini par sa transformée de Fourier : $\hat{H} = |\hat{h}|^2$.

Le problème (4) peut être interprété comme un problème d'attraction-répulsion [9] : des particules chargées positivement sont attirées par les zones claires de l'image et elles se repoussent deux à deux.

Algorithmie Bien que le problème (4) soit fortement non convexe, il peut être résolu par des techniques simples de type descente de gradient [3]. Il y a cependant trois difficultés techniques importantes : le choix du noyau H , le calcul des intégrales et le calcul du gradient de la fonctionnelle. Le noyau H doit être choisi de telle manière que des particules éloignées interagissent ensemble, ce qui semble assurer empiriquement une convergence assez rapide des algorithmes de descente de gradient. En pratique, nous avons choisi $H(x) = \|x\|_2$ dans ce papier. Pour ce choix, on obtient formellement :

$$[\nabla F(\mathbf{p})][i] = \frac{1}{n^2} \sum_{j \neq i} \frac{p_i - p_j}{\|p_i - p_j\|}, \quad (5)$$

et on voit que l'intensité de la force de répulsion de la particule j sur la particule i est *indépendante* de leur distance. Les intégrales peuvent être calculées avec des formules de quadrature. Enfin, lorsque le nombre de particules est grand, calculer le gradient (5) directement coûte $O(n^2)$ opérations. En pratique, on fait appel à des outils de simulation de particules tels que les méthodes multipolaires ou les transformées de Fourier non uniformes [8] qui permettent de réduire le coût par itération à $O(n \log(n/\epsilon))$ opérations où ϵ représente la précision désirée sur le calcul de sommations rapide.

2.2 W_2 discret

Théorie On suppose ici que μ est une mesure atomique de la forme

$$\mu = \sum_{j=1}^k \mu_j \delta_{m_j}. \quad (6)$$

où les points m_j appartiennent à une grille cartésienne. Ainsi, l'image à quantifier est déjà discrète. Dans ces conditions, le plan de transport γ^* est discret, c'est-à-dire qu'il s'écrit sous la forme $\gamma^* = \sum_{ij} \gamma_{ij} \delta_{p_i \times m_j}$. Soit C la matrice de coût définie par $C_{ij} = \|p_i - m_j\|_2^2$, on obtient $d_{W_2}(\mu, \nu) = \langle C, \gamma^* \rangle$ où $\langle \cdot, \cdot \rangle$ est le produit scalaire de Frobenius. Le plan de transport optimal, pour \mathbf{p} fixé, est donné par la solution du problème linéaire :

$$\gamma^*(\mathbf{p}) = \underset{\substack{\gamma \geq 0 \\ \sum_i \gamma_{ij} = \mu_j \\ \sum_j \gamma_{ij} = 1/n}}{\text{argmin}} \langle C, \gamma \rangle, \quad (7)$$

et (3) se réécrit comme :

$$d_{W_2}(\mu, \nu(\mathbf{p})) = \sum_{i=1}^n \sum_{j=1}^k \gamma_{ij}^*(\mathbf{p}) \|p_i - m_j\|_2^2. \quad (8)$$

À γ^* fixé, la minimisation de (8) donne

$$p_i = \frac{1}{n} \sum_j \gamma_{ij}^* m_j, \quad (9)$$

ce qui correspond à choisir la position p_i comme le barycentre des positions m_j pondérées par le plan de transport γ^* .

Pour résoudre (1) on propose d'utiliser une minimisation alternée en γ et en \mathbf{p} , à l'itération k , $\gamma^*(\mathbf{p}_k)$ est obtenu en résolvant (7) avec un solveur linéaire, puis \mathbf{p}_{k+1} est obtenu par la relation (9).

Algorithmie Le calcul de la distance de Wasserstein est un problème de programmation linéaire (PL) à nk variables. Beaucoup de solveurs permettent de le faire, mais ils sont trop coûteux pour les dimensions typiques rencontrées en traitement du signal (e.g. $k = 10^6$ et $n = 10^5$). Ce type de méthode ne peut donc être appliqué qu'en dimension relativement modérée. De plus, les solutions de (7) ne sont pas continues par rapport à la variable \mathbf{p} . Ce phénomène peut entraîner des instabilités numériques lors de l'optimisation.

2.3 W_2 discret régularisé

Théorie Quand le nombre de points de discrétisation ($n+k$) est grand, ou quand la mesure cible μ est bruitée, le plan de transport donné par (7) peut-être très irrégulier. Pour régulariser le problème, une méthode popularisée récemment consiste à ajouter un terme entropique [4, 2]. Soit

$$E(\gamma) = \sum_{ij} \gamma_{ij} (\log(\gamma_{ij}) - 1) \quad (10)$$

l'entropie de γ . Dans ces méthodes, on remplace $\langle C, \gamma \rangle$ dans (7) par $\langle C, \gamma \rangle - \varepsilon E(\gamma)$. Le problème ε -régularisé devient alors strictement convexe et admet une unique solution, il peut être réécrit sous la forme :

$$W_{2,\varepsilon}(\mu, \nu) = \min_{\gamma \in \Pi(\nu, \mu)} \mathbf{KL}(\gamma | \xi(\varepsilon)) \text{ avec } \xi(\varepsilon) = e^{-\frac{C}{\varepsilon}}, \quad (11)$$

où $\mathbf{KL}(\gamma | \xi) = \sum_{ij} \gamma_{ij} (\log(\gamma_{ij}) - \log(\xi_{ij}) - 1)$ est la divergence de Kullback-Leibler.

Algorithmie Une méthode de résolution numérique popularisée récemment [4, 2], consiste à utiliser des algorithmes simples de projections itérées de Bregman, voir algorithme 1. Dans cet algorithme l'exponentielle de matrice, la division et la multiplication entre vecteurs sont des opérations terme à terme. Le vecteur $\boldsymbol{\mu}$ représente les poids de la mesure cible. Le nombre de variables d'optimisation du problème direct est limité ($n+k$) mais la matrice ξ contient nk éléments et ne doit en général pas être calculée. Les produits $\xi \mathbf{u}$ et $\xi \mathbf{v}$ peuvent être calculés

Data : $C, \varepsilon, \mu, \text{IterMax}$

Result : γ^* le plan de transport régularisé

Initialisation : $\mathbf{v} = \mathbf{1}, \xi = e^{-\frac{C}{\varepsilon}};$

for $l \leq \text{nit} \leq \text{IterMax}$ **do**

$$\begin{cases} \mathbf{u} = \frac{\mathbf{1}}{n\xi\mathbf{v}}; \\ \mathbf{v} = \frac{\mu}{\xi^T\mathbf{u}}; \end{cases}$$

end

return $\gamma^* = \text{Diag}(\mathbf{u})\xi\text{Diag}(\mathbf{v})$

Algorithme 1 : Algorithme de Sinkhorn

efficacement avec la transformée de Fourier rapide si les points \mathbf{m} sont sur une grille cartésienne.

L'algorithme 1 fournit en général une solution approchée acceptable en une centaine d'itérations. Le coût de chaque itération est dominé par les produit $\xi\mathbf{v}$ et $\xi^T\mathbf{u}$ qui peuvent être effectués en $O(n \log(n))$ opérations.

Une fois γ^* obtenu par l'algorithme présenté précédemment, on utilise la relation (8) pour optimiser la position des points.

Le choix du paramètre de régularisation ε est très délicat. Si ε est trop grand, les masses de Dirac s'agglutinent (voir Figure 2). A l'inverse si ε est trop petit les vecteurs $u = \frac{\nu}{\xi\mathbf{v}}$ et $v = \frac{\mu}{\xi^T\mathbf{u}}$ deviennent trop grand (ce qui arrive rapidement comme $\xi = e^{-\frac{C}{\varepsilon}}$) et ne peuvent pas être représentés en double précision. Globalement, on voit qu'une implémentation "naïve" de cette méthode n'est pas satisfaisante.

2.4 W_2 semi-discret

Théorie Le cas semi-discret consiste à ne pas décrire μ comme une somme de masses de Dirac, mais comme une mesure à densité continue. Par exemple, on peut trianguler les pixels de l'image et supposer que l'image est linéaire sur chaque triangle. Un grand avantage de cette formulation est que le plan de transport est unique [12], ce qui assure une certaine stabilité. La forme duale du problème (3) est donnée par [12] :

$$W_2(\mu, \nu) = \sup_{\substack{\phi \in L^1(\Omega), \\ \psi \in L^1(\Omega), \\ \text{s.c. } \phi(x) + \psi(y) \leq \|x - y\|_2^2}} \int_{\Omega} \phi(x) d\mu(x) + \int_{\Omega} \psi(y) d\nu(y), \quad (12)$$

La résolution exacte de ce problème selon la variable ϕ donne l'équation suivante :

$$W_2(\mu, \nu) = \sup_{\psi} \int_{\Omega} \min_i \|x - p_i\|_2^2 - \psi(p_i) d\mu + \frac{1}{n} \sum_i \psi(p_i). \quad (13)$$

Ce problème peut être résolu numériquement en introduisant les cellules de Laguerre définies par :

$$\mathcal{L}_i = \{x \in \Omega, \|x - p_i\|_2^2 - w_i \leq \|x - p_j\|_2^2 - w_j, \forall j \neq i\}.$$

Le diagramme de Laguerre $(\mathcal{L}_i)_{1 \leq i \leq n}$ est l'union de n polygones convexes formant une partition de Ω [1]. En posant $\psi(p_i) = w_i$, le problème (12) se réécrit comme la maximisation de la fonctionnelle $g(w)$ [7], avec

$$g(w) = \sum_i \left(\int_{\mathcal{L}_i} (\|x - p_i\|_2^2 - w_i) \right) d\mu + \frac{1}{n} \sum_i w_i. \quad (14)$$

où $g(w)$ est une fonction concave, sa dérivée et sa Hessienne sont données dans [6].

Algorithme La solution w^* de (14) pour \mathbf{p} fixé peut être calculée avec un algorithme de Newton avec recherche linéaire de Wolfe [6]. A chaque évaluation du poids w^* , la position des points \mathbf{p} peut être optimisée grâce à un algorithme de type Lloyd : chaque masse de Dirac est déplacée sur le barycentre de la cellule de Laguerre associée :

$$p_i = \frac{\int_{\mathcal{L}_i} x d\mu(x)}{\mu(\mathcal{L}_i)}. \quad (15)$$

Dans notre implémentation, le calcul du diagramme de Laguerre est réalisé par la librairie CGAL [11]. Celui-ci peut être calculé en $O(n \log(n))$ opérations. Une fois les cellules de Laguerre (polygones) calculées, les intégrations nécessaires au calcul des dérivées de $g(w)$ sont réalisées rapidement grâce à la formule de Green (intégration de fonctions sur les arêtes uniquement). On peut montrer que le coût de calcul est en $O(\sqrt{kn})$ opérations. Cet algorithme couple une méthode de second ordre pour le calcul du transport optimal et une méthode de premier ordre pour l'optimisation de la position des points. Il fournit des résultats satisfaisants visuellement en très peu d'itérations. Par exemple, les auteurs de [5] suggèrent que quatre itérations suffisent en général.

3 Comparaison des méthodes

La Figure 1 montre un premier test de quantification de l'indicatrice d'un disque. La Figure 2 montre l'influence du paramètre de régularisation ε introduit en Section 2.3. La Figure 3 montre un résultat de stippling sur un dégradé linéaire. Ces quelques résultats montrent que les méthodes les plus simples à implémenter (transport optimal discret et transport optimal régularisé) souffrent malheureusement de sérieux problèmes numériques qui les rendent inutilisables pour le problème de stippling. Le transport optimal semi-discret et la distance d_{L^2} fournissent tous les deux des résultats très satisfaisants, cependant l'implémentation est bien plus difficile. En terme de rapidité de calcul, le transport optimal semi-discret semble le meilleur. Cependant il n'est pas clair que ce constat reste valable en dimension supérieure, car la complexité de construction des diagrammes de Laguerre augmente, alors que la complexité de la NFFT est inchangée. Pour finir le tableau 4 donne quelques propriétés supplémentaires des différentes méthodes.

Références

- [1] F. Aurenhammer, F. Hoffmann, and B. Aronov. Minkowski-type theorems and least-squares clustering. *Algorithmica*, 20(1) :61–76, 1998.
- [2] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regulari-

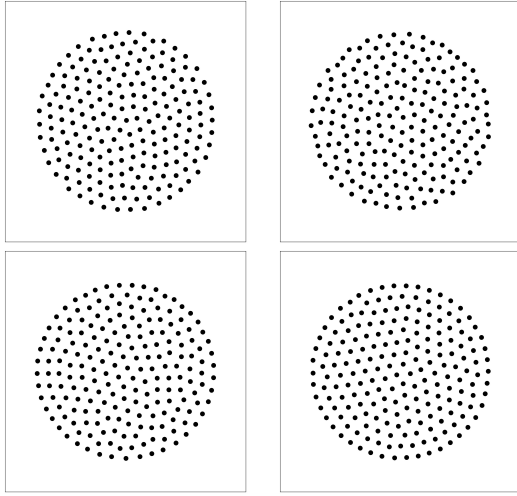


FIGURE 1 – Approximation de la fonction caractéristique d’un disque par les méthodes : L^2 (en haut à gauche), W_2 discret (en haut à droite), W_2 discret régularisé (en bas à gauche) et W_2 semi-discret (en bas à droite). Ici, toutes les méthodes produisent des résultats satisfaisants. Noter cependant les nombreuses irrégularités obtenues avec le transport optimal discret.

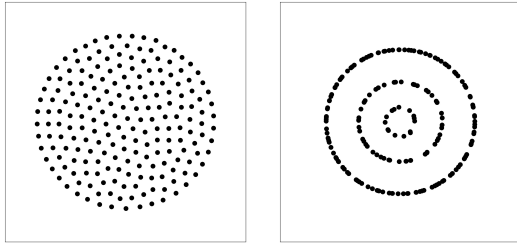


FIGURE 2 – Approximation de la fonction caractéristique d’un disque par la méthode de transport optimal régularisé pour un paramètre de régularisation ε valant respectivement $2.5 \cdot 10^{-4}$ et $6.0 \cdot 10^{-3}$. Si ε est trop important, la régularisation entropique a tendance à agréger les points plutôt que les disposer de façon uniforme.

zed transportation problems. *SIAM Journal on Scientific Computing*, 37(2) :A1111–A1138, 2015.

- [3] N. Chauffert, P. Ciucu, J. Kahn, and P. Weiss. A projection method on measures sets. *Constructive Approximation*, 45(1) :83–111, 2017.
- [4] M. Cuturi. Sinkhorn distances : Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.
- [5] F. De Goes, K. Breeden, V. Ostromoukhov, and M. Desbrun. Blue noise through optimal transport. *ACM Transactions on Graphics (TOG)*, 31(6) :171, 2012.
- [6] J. Kitagawa, Q. Mérigot, and B. Thibert. A Newton algorithm for semi-discrete optimal transport. *arXiv preprint arXiv :1603.05579*, 2016.

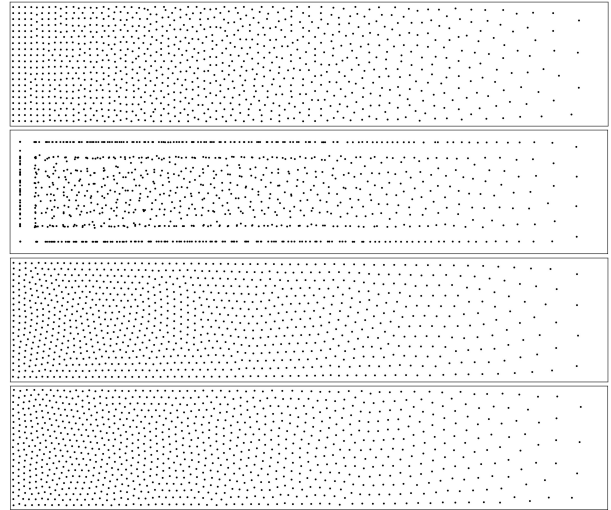


FIGURE 3 – Stipling d’un dégradé linéaire. De haut en bas : W_2 discret, W_2 discret régularisé, W_2 semi-discret et d_{L^2} . Noter que le transport optimal discret tend à aligner les points avec la grille, car les masses de Dirac de la mesure cible sont localisées sur une grille. Pour le transport régularisé, la valeur de ε est la plus petite possible avant que des NaN apparaissent, car des divisions par des nombres trop petits ont lieu.

Méthode	Propriétés					
	Invariance isométrie	Dim. supérieure	Unicité γ^*	Robustesse bruit	Bons résultats	Technologie
d_{L^2} (semi-discret)	✗	✓	✓	✓	✓	NFFT
d_{W_2} discret	✓	✓	✗	✗	✗	PL
d_{W_2} régularisé	✓	✓	✓	✓	✗	FFT
d_{W_2} semi-discret	✓	✗	✓	✓	✓	CGAL

FIGURE 4 – Quelques propriétés des différents algorithmes.

- [7] B. Lévy. A numerical algorithm for 12 semi-discrete optimal transport in 3d. *ESAIM : Mathematical Modelling and Numerical Analysis*, 49(6) :1693–1715, 2015.
- [8] D. Potts and G. Steidl. Fast summation at nonequispaced knots by nfft. *SIAM Journal on Scientific Computing*, 24(6) :2013–2037, 2003.
- [9] C. Schmalz, P. Gwosdek, A. Bruhn, and J. Weickert. Electrostatic halftoning. In *Computer Graphics Forum*, volume 29, pages 2313–2327. Wiley Online Library, 2010.
- [10] S. Smale. Mathematical problems for the next century. *The Mathematical Intelligencer*, 20(2) :7–15, 1998.
- [11] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 4.9 edition, 2016.
- [12] C. Villani. *Optimal transport : old and new*, volume 338. Springer Science & Business Media, 2008.