Initialization Procedures for Discrete and Semi-Discrete Optimal Transport

Jocelyn Meyron

Version of Record: https://www.sciencedirect.com/science/article/pii/S0010448519301770
Manuscript_910d44bc42d683f68c0a4eebe5ae5abb

Initialization procedures for discrete and semi-discrete optimal transport

Jocelyn Meyron*

INSA Lyon, LIRIS, CNRS, Lyon, France

Abstract

We propose three different methods to initialize optimal transport algorithms in both the discrete and semi-discrete settings. After introducing the optimal transport problem, we start by explaining why finding "good" (in a certain sense) initial weights is an important problem in computational optimal transport. We then describe three novel procedures to find such weights. Proofs of correctness are also given. We finally show on many numerical examples how choosing these weights improve the running times of optimal transport algorithms. We also describe some applications in various fields such as non-imaging optics; matching between a point cloud and a triangulated surface; seismic imaging.

Keywords: optimal transport, Laguerre diagram, computational geometry, Newton method, entropic regularization

1. Introduction

Recently, optimal transport has become a useful tool in mathematics from both a theoretical and numerical point of view due to its connections with many different fields such as image processing [1], machine learning [2, 3, 4] and the study of partial differential equations [5]. Efficient numerical methods have also been proposed based on the different formulations of optimal transport. For instance, we will see in the next section that, in some settings, it can be formulated as an optimization problem and thus be solved using traditional methods such as Newton-based algorithms. It is also well known that the choice of the starting point in such methods is very important as it directly impacts its convergence properties. In this paper, we are interested in this aspect of optimal transport that is the choice of the starting point in different numerical methods.

We start by a brief introduction on optimal transport and we detail the existing numerical methods for solving it in two different settings. We also discuss why the choice of the starting point is important and even mandatory in some practical settings. We then present three original methods for finding such weights before using them on different numerical examples and applications.

1.1. Optimal transport

We introduce here the basic notions of optimal transport, for more details on both its theoretical and numerical aspects, one can refer to [6, 7, 8, 9] and references therein.

The optimal transport is an old problem in mathematics dating back to the 18th century that was first stated by *Gaspard Monge* [10]. One is interested in *transporting* a probability measure μ supported on a set $X \subset \mathbb{R}^d$ onto another probability measure ν supported on a set $Y \subset \mathbb{R}^d$ while minimizing a given cost function $c : X \times Y \to \mathbb{R}$. More precisely the so-called *Monge formulation* can be expressed as the following minimization problem:

$$\min_T \left\{ \int_X c(x, T(x)) \mathrm{d}\mu(x) \mid T \text{ transport map between } X \text{ and } Y \right\}$$

where a *transport map* T is a map from X to Y that preserves the mass i.e. $\forall B \subset Y,\ \mu(T^{-1}(B)) = \nu(B)$.

In the 20th century, Kantorovich [11] proposed another formulation that allows mass to split (which is not possible in Monge formulation since T must be a valid map between X and Y). The idea is to replace the transport map T by a *transport plan* γ which is a probability measure on the product space $X \times Y$. The problem becomes:

$$\min_\gamma \left\{ \int_{X\times Y} c(x, y) \mathrm{d}\gamma(x, y) \mid \gamma \text{ transport plan } X \text{ and } Y \right\}$$

where a *transport plan* γ is a probability measure on $X \times Y$ with marginals μ and ν. With this formulation, the problem becomes a linear programming problem with convex constraints. One can then look at its dual formulation, which can be easily obtained and is the following:

$$\max_{\varphi,\psi} \left\{ \int_X \varphi(x) \mathrm{d}\mu(x) - \int_Y \psi(y) \mathrm{d}\nu(y) \mid \forall x, y,\ \varphi(x) - \varphi(y) \leqslant c(x, y) \right\}.$$

Under some assumptions on the cost function and the geometry of X and Y, one can show that the two problems have solutions that coincide [7]. The solutions of this dual problem (φ, ψ) are called *Kantorovich potentials*.

1.2. Computational optimal transport

We now detail in two settings the numerical methods that exist to solve optimal transport problems.

Discrete setting

When both $\mu = \sum_{x\in X} \mu_x \delta_x$ and $\nu = \sum_{y\in Y} \nu_y \delta_y$ are discrete probability measures i.e. supported on point clouds in $\mathbb{R}^d$, efficient numerical methods exist. Most of them are based on the so-called *entropic regularization* of optimal transport [2] that is a relaxation of the dual Kantorovich formulation. It can be stated as the following unconstrained maximization problem:

$$\max_{\varphi,\psi} \sum_{(x,y)\in X\times Y} \epsilon \exp\left(\frac{1}{\epsilon}(\varphi(x) - \psi(y) - c(x, y))\right) + \sum_{x\in X} \varphi(x)\mu_x - \sum_{y\in Y} \psi(y)\nu_y$$

*Corresponding author
Email address: Jocelyn.Meyron@insa-lyon.fr (Jocelyn Meyron)

where $\epsilon > 0$ is a regularization parameter. The main numerical methods available to solve this problem are variations of the so-called Sinkhorn-Knopp algorithm, see Algorithm 1 for its simplest version. This algorithm is interesting since it is extremely easy to implement as it only relies on matrix multiplication. In this algorithm, $\mathbf{1}_X$ denotes the vector with only ones with size the number of elements in X. G_ϵ is sometimes called the *Gibbs matrix*. Because of its ease of implementation, it has been used with success in numerous applications such as computing barycenters of measures [12] or surface matching [13].

Input Measures μ, ν, cost matrix $C = [c(x, y)]_{x\in X, y\in Y}$
Regularization parameter $\epsilon > 0$
Maximum number of iterations $k_{max} > 0$
Output Kantorovich potentials φ, ψ solving the optimal transport problem between μ and ν for the cost c
Set $u_0 \leftarrow \mathbf{1}_X$, $v_0 \leftarrow \mathbf{1}_Y$ and $G_\epsilon = \exp(-\frac{C}{\epsilon})$
For $k = 0$ to k_{max}

- $u_{k+1} \leftarrow \mu/(G_\epsilon v_k)$
- $v_{k+1} \leftarrow \nu/(G_\epsilon^T u_{k+1})$

Set $\varphi = \epsilon \ln(u_{k_{max}})$ and $\psi = -\epsilon \ln(v_{k_{max}})$
Algorithm 1: Sinkhorn-Knopp algorithm

Semi-discrete setting

We now look at the so-called *semi-discrete* setting where μ is continuous and ν is a discrete probability measure. In this case, efficient numerical methods relying on tools from computational geometry [14], namely *Power diagrams*, have been developed. Their convergence is well studied and make them tractable for real-life applications [15, 16]. We briefly detail here the main ingredients of this formulation. We denote $\nu = \sum_{i=1}^N \nu_i \delta_{y_i}$ the target probability measure. For a vector $\psi \in \mathbb{R}^N$, we define the *Laguerre cell* of y_i by

$$\mathrm{Lag}_i(\psi) = \{x \in X \mid \forall j,\ c(x, y_i) + \psi_i \leqslant c(x, y_j) + y_j\}.$$

One can then show that the *Laguerre diagram* (the collection of all the Laguerre cells) is a partition of X and that the dual Kantorovich formulation is equivalent to maximizing the following function Φ:

$$\Phi(\psi) = \sum_{i=1}^N \int_{\mathrm{Lag}_i(\psi)} (c(x, y_i) + \psi_i) \mathrm{d}\mu(x) - \sum_{i=1}^N \nu_i \psi_i.$$

Under some assumptions on c and the geometry of the supports of the source and target measures, see [17, 18, 19, 20], Φ is shown to be concave and of class C^2. Its gradient is given by $\nabla\Phi(\psi) = G(\psi) - \nu$ where $G(\psi) = (G_i(\psi))_{1\leqslant i\leqslant N} = (\mu(\mathrm{Lag}_i(\psi))_{1\leqslant i\leqslant N}$ is the vector of the areas of the Laguerre cells for a weight vector ψ. Thus maximizing Φ means that we need to

$$\text{Find a vector } \psi \in \mathbb{R}^N \text{ such that } G(\psi) = \nu. \qquad \text{(DMA)}$$

This equation can be seen as a discretization of the so-called *Monge-Ampère* equation hence its name (DMA). To solve this equation, multiple numerical methods have been proposed. The most interesting one is the so-called *damped Newton's method* [21, 17] since its convergence speed has been studied in different settings, making it useful for many applications. The details of the method can be found in Algorithm 2. In this algorithm, $\mathrm{D}G^+$ denotes the *pseudo-inverse* of the Jacobian matrix $\mathrm{D}G$. This method is a variation of the Newton's method with the added condition that it preserves the fact that all the Laguerre cells have positive mass (condition $\epsilon_0 > 0$ at the start and the second item in the loop).

Input A source measure μ
A target measure $\nu = \sum_{1\leqslant i\leqslant N} \nu_i \delta_{y_i}$
A numerical error $\eta > 0$
A family of weights $\psi^0 \in \mathbb{R}^N$ such that $\varepsilon_0 := \min\left[\min_i G_i(\psi^0),\ \min_i \nu_i\right] > 0$

Output A family of weights ψ^k solving (DMA) up to η, i.e. $\left\|G(\psi^k) - \nu\right\| \leqslant \eta$.

While $\left\|G(\psi^k) - \nu\right\| \geqslant \eta$

- Compute $\nu^k = -\mathrm{D}G(\psi^k)^+(G(\psi^k) - \nu)$
- Determine the minimum $\ell \in \mathbb{N}$ such that $\psi^{k,\ell} := \psi^k + 2^{-\ell}\nu^k$ satisfies

$$\begin{cases} \min\limits_{1\leqslant i\leqslant N} G_i(\psi^{k,\ell}) \geqslant \varepsilon_0 \\ \left\|G(\psi^{k,\ell}) - \nu\right\| \leqslant (1 - 2^{-(\ell+1)}) \left\|G(\psi^k) - \nu\right\| \end{cases}$$

- Set $\psi^{k+1} = \psi^k + 2^{-\ell}\nu^k$ and $k \leftarrow k + 1$.

Algorithm 2: Damped Newton's algorithm

1.3. Initialization problem

In both settings, the choice of initial parameters, (u_0, v_0) in the discrete setting and ψ^0 in the semi-discrete one, is important. In the following, we will call this choice the *initialization problem*. For instance, in the semi-discrete setting, for the damped Newton's method to converge, see [18], a necessary condition is that all the Laguerre cells must have positive mass at every step of the algorithm and in particular at the beginning (stated by the condition $\varepsilon_0 > 0$). We will see in Section 3 that this condition can be hard to ensure in practice. In the discrete setting, since there is no notion of Laguerre cells, this problem does not arise but we will show in Section 3 that choosing "better" initial vectors (u_0, v_0) will accelerate the convergence of Algorithm 1.

Not a lot of work was dedicated to the study of this problem. In particular, since in the discrete setting, there is no a priori constraint on the choice of initial parameters, the need of better initial vectors is not obvious. The natural question would be: can we improve the running time by choosing "better" initial parameters? We will see that numerical experiments answer this question positively. In the semi-discrete setting, in most of the applications, the target points lie in the source domain avoiding this initialization issue. However, we will see in Section 3 that there are some settings where this is not the case anymore and where we need to also have efficient initialization procedures.

2. Initialization procedures

In this section, we present three initialization procedures that can be used to compute initial parameters for speeding up optimal transport algorithms in both the discrete and semi-discrete settings. We also provide proofs of their correctness. We will also detail in which specific settings each method works the best. We start by looking at the semi-discrete setting.

2.1. Local perturbation

The first method we will present only works for the quadratic cost function $c(x,y) = \|x-y\|^2$ and when the support of the source measure X is a compact set. It relies on the following proposition (presented in [19] and included here for completeness) that describes a way to choose weights such that all the Laguerre cells are not empty.

Proposition 1. *Let $X \subset \mathbb{R}^d$ be a compact set, $Y = \{y_1, \dots, y_N\} \subset \mathbb{R}^d$ be a point set and $\psi_i^0 = -d_X(y_i)^2$. Then*

$$\emptyset \neq \{x \in X \mid d_X(y_i) = \|x - y_i\|\} \subset \mathrm{Lag}_i(\psi^0)$$

where $d_X(y_i) = \min_{x\in X} \|x - y_i\|$.

Proof. Let $i \in \{1, \dots, N\}$ and $x \in X$ such that $d_X(y_i) = \|x - y_i\|$, then for $j \in \{1, \dots, N\}$

$$\begin{aligned}\left\|x - y_j\right\|^2 + \psi_j^0 &= \left\|x - y_j\right\|^2 - d_X(y_j)^2 \\ &\geqslant d_X(y_j)^2 - d_X(y_j)^2 = 0 = \|x - y_i\|^2 + \psi_i^0.\end{aligned}$$

Thus $x \in \mathrm{Lag}_i(\psi^0)$. □

We can then iteratively perturb these weights to make the Laguerre cells have positive mass, this is the result of the next proposition. It makes use of assumptions on the source and target measures namely that μ should be a *regular simplicial measure* (a measure supported on a triangulated surface in $\mathbb{R}^3$ for instance) and the support of ν should be in *generic position* with respect to the support of μ. More details about these assumptions can be found in [19] such as the proof of the convergence of the damped Newton's method in this specific setting. In all the numerical experiments, we will choose X to be a triangulated surface in $\mathbb{R}^3$, setting in which the convergence result applies. We denote by $Z(\psi)$ the following set:

$$Z(\psi) = \{i \in \{1, \dots, N\} \mid G_i(\psi) = 0\}$$

i.e. the set of indices of Laguerre cells with zero mass. We also define the *Power cell* of y_i

$$\mathrm{Pow}_i(\psi) = \{x \in \mathbb{R}^d \mid \forall j,\ \|x - y_i\|^2 + \psi_i \leqslant \left\|x - y_j\right\|^2 + \psi_j\}.$$

Let us remark that $\mathrm{Lag}_i(\psi) = X \cap \mathrm{Pow}_i(\psi)$.

Proposition 2. *Let μ be a regular simplicial measure supported on X, Y be a point cloud in generic position with respect to X. Let $\psi \in \mathbb{R}^N$ be a vector such that $Z(\psi) \neq \emptyset$ and for every $i \in \{1, \dots, N\}$, $\mathrm{Pow}_i(\psi) \cap X \neq \emptyset$. Then there exists $\epsilon_0 > 0$ such that for $\epsilon < \epsilon_0$*

$\mathrm{Card}(Z(\widetilde{\psi})) < \mathrm{Card}(Z(\psi))$ *and* $\forall i \in \{1, \dots, N\},\ \mathrm{Pow}_i(\widetilde{\psi}) \cap X \neq \emptyset$

where $\widetilde{\psi} = \psi - \epsilon \mathbf{1}_{Z(\psi)}$. $\mathbf{1}_A$ denotes the vector of $\mathbb{R}^N$ with a one on the ith row if $i \in A$ and a zero otherwise.

For the proof, we define, for $i \neq j$, the halfspace $H_{i,j}(\psi)$ by

$$H_{i,j}(\psi) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 + \psi_i \leqslant \left\|x - y_j\right\|^2 + \psi_j\}.$$

Let us remark that we have $\mathrm{Pow}_i(\psi) \subset H_{i,j}(\psi)$ for any $j \neq i$. We will also need the two following lemmas.

Lemma 3. *Let $i \in \{1, \dots, N\}$, $\epsilon > 0$, and $\widetilde{\psi} = \psi - \epsilon \mathbf{1}_{\{i\}}$, then the distance between $\partial H_{i,j}(\psi)$ and $\partial H_{i,j}(\widetilde{\psi})$ is given by*

$$d(\partial H_{i,j}(\psi), \partial H_{i,j}(\widetilde{\psi})) = \frac{\epsilon}{2\left\|y_i - y_j\right\|}.$$

Let us remark that $H_{i,j}(\psi) \subset H_{i,j}(\widetilde{\psi})$ meaning that $\partial H_{i,j}(\widetilde{\psi})$ moves closer to y_j.

Lemma 4. *Let μ be a regular simplicial measure supported on X, p a point in X and $\psi \in \mathbb{R}^N$, we define the set of empty cells containing p by*

$$Z_p(\psi) = \{i \in Z(\psi) \text{ and } p \in \mathrm{Pow}_i(\psi) \cap X\}.$$

If $Z_p(\psi) \neq \emptyset$ and $\widetilde{\psi} = \psi - \epsilon \mathbf{1}_{Z(\psi)}$ for $\epsilon > 0$, there exists $r > 0$ such that

$$B(p, r) \subset \bigsqcup_{j \in Z_p(\psi)} \mathrm{Pow}_j(\widetilde{\psi})$$

where $B(p, r)$ denotes the ball of center p and radius r and $\bigsqcup$ the disjoint union.

Proof. We take a point $p \in X$, a weight vector $\psi \in \mathbb{R}^N$, $i \in Z_p(\psi)$, $\epsilon > 0$ and define $\widetilde{\psi} = \psi - \epsilon \mathbf{1}_{Z(\psi)}$. We also take $j \notin Z_p(\psi)$.

By definition, we have $\mathrm{Pow}_i(\psi) = \bigcap_{k\neq i} H_{i,k}(\psi)$, thus Lemma 3 implies that if we choose $r < \frac{\epsilon}{2\|y_i - y_j\|}$ then $B(p, r) \subset H_{i,j}(\widetilde{\psi})$. Furthermore, if we choose

$$r = \frac{\epsilon}{4 \min\limits_{i \in Z_p(\psi), j \notin Z_p(\psi)} \left\|y_i - y_j\right\|},$$

then we get

$$\forall\, j \notin Z_p(\psi),\ B(p, r) \subset \bigcup_{i \in Z_p(\psi)} H_{i,j}(\widetilde{\psi}).$$

Taking the complement that we denote by $X^{\complement}$ (for a set $X \subset \mathbb{R}^d$), we get

$$\begin{aligned}\forall\, j \notin Z_p(\psi),\ B(p, r)^{\complement} \supset \left(\bigcup_{i \in Z_p(\psi)} H_{i,j}(\widetilde{\psi})\right)^{\complement} &= \bigcap_{i \in Z_p(\psi)} H_{i,j}(\widetilde{\psi})^{\complement} \\ &= \bigcap_{i \in Z_p(\psi)} H_{j,i}(\widetilde{\psi}) \supset \mathrm{Pow}_j(\widetilde{\psi}).\end{aligned}$$

This means

$$\bigcap_{j \notin Z_p(\psi)} \mathrm{Pow}_j(\widetilde{\psi}) \subset B(p, r)^{\complement}.$$

Finally, taking the complement again and using the fact that the Power diagram is a partition of $\mathbb{R}^d$, we obtain the intended result i.e.

$$B(p, r) \subset \bigcup_{j \notin Z_p(\psi)} \mathrm{Pow}_j(\widetilde{\psi})^{\complement} = \bigcup_{j \in Z_p(\psi)} \mathrm{Pow}_j(\widetilde{\psi}).$$

□

Proof of Proposition 2. We take $\psi \in \mathbb{R}^N$ such that $Z(\psi) \neq \emptyset$ and for every i, $\mathrm{Pow}_i(\psi) \cap X \neq \emptyset$. According to Theorem 14 in [19], G is of class C^1. In particular it is continuous, so we can find $\epsilon_0 > 0$ such that for $\epsilon < \epsilon_0$ and $\widetilde{\psi} = \psi - \epsilon \mathbf{1}_{Z(\psi)}$, we have: $G_i(\psi) > 0 \implies G_i(\widetilde{\psi}) > 0$. Furthermore, if $i \in Z(\psi)$ then $\mathrm{Pow}_i(\psi) \subset \mathrm{Pow}_i(\widetilde{\psi})$. We conclude that if $\mathrm{Pow}_i(\psi) \cap X$ is not empty then $\mathrm{Pow}_i(\widetilde{\psi}) \cap X$ stays not empty.

We now take $p \in X$ such that $Z_p(\psi) \neq \emptyset$ and $\epsilon < \epsilon_0$, then Lemma 4 gives the existence of $r > 0$ such that

$$B(p, r) \subset \bigsqcup_{j \in Z_p(\psi)} \mathrm{Pow}_j(\widetilde{\psi}).$$

We know that p belongs to a simplex σ. Since μ is a regular simplicial measure, the density μ_σ with respect to the $\dim(\sigma)$-dimensional Hausdorff measure on σ is bounded from below (second item in the definition of a regular simplicial measure), so that $\mu(B(p,r) \cap \sigma) > 0$. Thus $0 < \mu(B(p,r) \cap \sigma) \leqslant \sum_{j \in Z_p(\psi)} \mu(\mathrm{Pow}_j(\widetilde{\psi}) \cap \sigma)$. This means that there exists $j \in Z_p(\psi)$ such that $\mu(\mathrm{Pow}_j(\widetilde{\psi}) \cap \sigma) > 0$ i.e. $j \notin Z(\widetilde{\psi})$. Thus we found a Laguerre cell $\mathrm{Lag}_j(\psi)$ that gained mass i.e. $\mathrm{Card}(Z(\widetilde{\psi})) < \mathrm{Card}(Z(\psi))$. □

We can then combine Proposition 1 and iteratively apply Proposition 2 to obtain the following result which guarantees that no Laguerre cell will have a mass of zero after the initialization step.

Proposition 5. *Let μ be a regular simplicial measure supported on $X \subset \mathbb{R}^d$, $Y \subset \mathbb{R}^d$ a point cloud in generic position with respect to X. Let us set $\psi_i^0 = -d_X(y_i)^2$ such that $\mathrm{Lag}_i(\psi^0) \neq \emptyset$ for all $i \in \{1, \dots, N\}$. Then there exists $\widetilde{\psi} \in \mathbb{R}^N$ such that $\mathrm{Card}(Z(\widetilde{\psi})) = 0$.*

Proof. Iterating Proposition 2 starting from the weights ψ^0, we obtain an integer sequence $(\mathrm{Card}(Z(\psi^i)))_{i \geqslant 0}$ that is strictly decreasing, thus converging towards 0. □

The detailed procedure corresponding to this proposition can be found in Algorithm 3. Its convergence is directly implied by Proposition 5. This method only needs one parameter that is ϵ i.e. the maximal quantity for which we decrease the weights ψ. What the inner loop does is checking whether for this choice of ϵ we obtain a smaller amount of Laguerre cells with zero-mass. If it is the case, we exit the loop. If not, we halve ϵ and try again. The previous proposition ensures that such a choice of ϵ exists (even if it is very small).

Input A regular simplicial measure μ
A finitely supported measure $\nu = \sum_{1 \leqslant i \leqslant N} \nu_i \delta_{y_i}$
A maximal decrement $\epsilon > 0$
A family of weights $\psi^0 \in \mathbb{R}^N$ such that $\forall i,\ \mathrm{Lag}_i(\psi^0) \neq \emptyset$

Output A family of weights ψ such that $\forall i,\ G_i(\psi) > 0$.

Initialization $\psi \leftarrow \psi^0, z \leftarrow \mathrm{Card}(Z(\psi^0))$

while $z > 0$ **do**
  $z_{cur} \leftarrow N$
  $\epsilon_{cur} \leftarrow \epsilon$
  while $z_{cur} > z$ **do**
    $\psi_{cur} \leftarrow \psi - \epsilon_{cur} \mathbf{1}_{Z(\psi_{cur})}$
    $z_{cur} \leftarrow \mathrm{Card}(Z(\psi_{cur}))$
    if $z_{cur} < z$ **then**
      $\psi \leftarrow \psi_{cur}$
      $z \leftarrow \mathrm{Card}(Z(\psi))$
      break
    else
      $\epsilon_{cur} \leftarrow \epsilon_{cur}/2$
    end
  end
end

Algorithm 3: Local perturbation method

2.2. Rescaling

In the second method, we make use of the relation between the Laguerre cells of a point cloud Y and the point cloud $Z = \lambda Y + t$ which is a uniform scaling followed by a uniform translation of Y, given by the following proposition. Let us note that this method also only works for the quadratic cost.

Proposition 6. *Given a point set $Y = \{y_1, \dots, y_N\} \subset \mathbb{R}^d$, $\lambda > 0$ and $t \in \mathbb{R}^d$, if we define the point set $Z = \{z_1, \dots, z_N\} \subset \mathbb{R}^d$ by $z_i = \lambda y_i + t$, then there is the following relation between the Laguerre cells of the two sets*

$$\mathrm{Lag}_{z_i}(\psi) = \mathrm{Lag}_{y_i}(\varphi)$$

where $\varphi_i = \frac{\psi_i}{\lambda} + 2\langle t \mid y_i \rangle + (\lambda - 1)\|y_i\|^2$ for $i \in \{1, \dots, N\}$.

Proof. Let us take $x \in \mathrm{Lag}_{z_i}(\psi)$ for a vector of weights $\psi \in \mathbb{R}^N$ and $j \in \{1, \dots, N\}$, we have

$$\begin{aligned}
\left\|z_j\right\|^2 - \|z_i\|^2 &= \langle z_j - z_i \mid z_j + z_i \rangle \\
&= \lambda \langle y_j - y_i \mid \lambda(y_j + y_i) + 2t \rangle \\
&= \lambda^2 (\left\|y_j\right\|^2 - \|y_i\|^2) + 2\lambda \langle t \mid y_j - y_i \rangle \\
&= \lambda \left[\left\|y_j\right\|^2 - \|y_i\|^2 + (\lambda - 1)(\left\|y_j\right\|^2 - \|y_i\|^2) + 2\langle t \mid y_j - y_i \rangle \right].
\end{aligned}$$

Thus

$$\begin{aligned}
x \in \mathrm{Lag}_{z_i}(\psi) &\iff \forall j,\ \|x - z_i\|^2 + \psi_i \leqslant \left\|x - z_j\right\|^2 + \psi_j \\
&\iff \forall j,\ -2\langle x \mid z_i - z_j \rangle \leqslant \psi_j - \psi_i + \left\|z_j\right\|^2 - \|z_i\|^2 \\
&\iff \forall j,\ -2\langle x \mid y_i - y_j \rangle \leqslant \varphi_j - \varphi_i + \left\|y_j\right\|^2 - \|y_i\|^2 \\
&\iff \forall j,\ \|x - y_i\|^2 + \varphi_i \leqslant \left\|x - y_j\right\|^2 + \varphi_j \\
&\iff x \in \mathrm{Lag}_{y_i}(\varphi).
\end{aligned}$$

□

There are no general ways for finding t and λ since they heavily depend on the geometry of X and Y. Two simple ways are presented below. We denote by $\overline{X}$ the *centroid* of a domain $X \subset \mathbb{R}^d$. If $X = \{x_1, \dots, x_N\}$ is finite then $\overline{X} = \frac{1}{N} \sum_{i=1}^N x_i$; if it is a continuous domain equipped with some measure μ then $\overline{X} = \frac{\int_X x \mathrm{d}\mu(x)}{\int_X \mathrm{d}\mu(x)}$. We also denote by $\mathrm{bbox}(X)$ an axis-aligned bounding box of a domain $X \subset \mathbb{R}^d$. Finally, the diameter of a compact set X is denoted by $\mathrm{diam}(X) = \max_{x,y \in X} \|x - y\|$. With these notations, we propose the following choices:

- $(t, \lambda) = \left(\overline{X} - \overline{Y}, \frac{\mathrm{diam}(X)}{\mathrm{diam}(Y)}\right)$,
- $(t, \lambda) = \left(\overline{\mathrm{bbox}(X)} - \overline{\mathrm{bbox}(Y)}, \frac{\mathrm{vol}(\mathrm{bbox}(X))}{\mathrm{vol}(\mathrm{bbox}(Y))}\right)$.

We then deduce the following initialization strategy.

Proposition 7. *Let μ be a probability measure supported on $X \subset \mathbb{R}^d$ and ν a discrete probability measure supported on a point set $Y \subset \mathbb{R}^d$. If there exists $t \in \mathbb{R}^d$ and $\lambda > 0$ such that $Z := \lambda Y + t \subset X$, then we can find $\psi \in \mathbb{R}^N$ such that $G_i(\psi) > 0$ for all $i \in \{1, \dots, N\}$.*

Proof. Let us suppose that there exists t and λ such that $Z := \lambda Y + t \subset X$. Then for the target measure $\nu_Z = \sum_{z_i \in Z} \nu_{y_i} \delta_{z_i}$ it suffices to set ψ^0 to a constant to ensure the positivity of the masses of the Laguerre cells (since $Z \subset X$) and we can solve the optimal transport problem between μ and ν_Z. We then use Proposition 6 to deduce optimal weights φ for the transport between μ and the original target measure ν i.e. we have $G_i(\varphi) = \nu_{y_i} > 0$. □

2.3. Linear interpolation

Finally, the last method we will introduce relies on a *linear interpolation* between two source measures: the original one μ and the normalized Lebesgue measure λ_P defined on a bigger domain P containing X and Y. The interpolated measure will be denoted by μ_t and is defined by

$$\forall t \in [0,1],\ \mu_t = t\lambda_P + (1-t)\mu.$$

The idea is to start with $t = 1$ i.e. $\mu_1 = \lambda_P$. Since P is chosen to be "big enough", one can choose weights ψ^0 for which all the Laguerre cells have positive mass. We then iteratively decrease the interpolation parameter t until we reach a sufficiently small value for which we are sure that the Laguerre cells for μ have positive mass. This minimal value for t is given by the next proposition:

Proposition 8. *For $t < \min_i \nu_i - \eta$ where η is the numerical error of Algorithm 2 then for every $i \in \{1,\dots,N\}$, $\mu(\mathrm{Lag}_i(\psi)) > 0$.*

For the proof, we will need the next lemma.

Lemma 9. *Let μ be a probability measure defined on X, λ_P the normalized Lebesgue measure on P and $\mu_t = t\lambda_P + (1-t)\mu$, for $t \in]0,1[$ be a probability measure on P. Then*

$$\forall A \subset P,\ \mu_t(A) > t \implies \mu(A) > 0.$$

Proof. We take $0 < t < 1$ and suppose that $\mu_t(A) > t$ and $\mu(A) = 0$. Then since $\mu_t(A) = t\lambda_P(A)$, we get $\lambda_P(A) > 1$ which is not possible since λ_P is a probability measure over P. Thus $\mu(A) > 0$. □

Proof of Proposition 8. For a numerical error η and $i \in \{1,\dots,N\}$, at the end of the optimal transport algorithm (see Algorithm 2) between two measures μ_t and $\nu = \sum_{i=1}^N \nu_i \delta_{y_i}$, we have $|\mu_t(\mathrm{Lag}_i(\psi)) - \nu_i| \leqslant \eta$. Thus $\mu_t(\mathrm{Lag}_i(\psi)) \geqslant \nu_i - \eta \geqslant \min_i \nu_i - \eta$. So if $t < \min_i \nu_i - \eta$ then $\mu_t(\mathrm{Lag}_i(\psi)) > t$ and we can apply Lemma 9 with $A = \mathrm{Lag}_i(\psi)$, so that $\mu(\mathrm{Lag}_i(\psi)) > 0$. □

Let us note that this method is not restricted to the quadratic cost and can work with any cost function. This can be useful in some applications, for instance in optical component design, see [22] and Section 3 for more details.

The details for this method can be found in Algorithm 4. In this algorithm $\texttt{SOLVE\_OT}(\mu, \nu, \eta, \psi^0)$ denotes a function that solves optimal transport between μ and ν, for a numerical error $\eta > 0$ and starting from weights ψ^0 (which can be Algorithm 2 for instance).

Input A source measure μ supported on X
A target measure $\nu = \sum_{1 \leqslant i \leqslant N} \nu_i \delta_{y_i}$
A set P that contains $X \cup Y$
A tolerance $\eta > 0$
Weights $\psi^0 \in \mathbb{R}^N$ such that $\forall i,\ \lambda_P(\mathrm{Lag}_i(\psi^0)) > 0$.

Output A family of weights ψ such that $\forall i,\ G_i(\psi) > 0$.

Initialization $k := 0$ and $t := 1$

while $t > t_{min} := \max(\min_i \nu_i - \eta, 0)$ **do**
  Define $\mu_t = t\lambda_P + (1-t)\mu$
  $\psi^{k+1} \leftarrow \texttt{SOLVE\_OT}(\mu_t, \nu, \eta, \psi^k)$
  $t \leftarrow t/2$
  $k \leftarrow k+1$
end

Algorithm 4: Interpolation method

2.4. Discrete setting

We now look at the discrete setting. We recall the relation between Kantorovich potentials and Sinkhorn vectors that is $(\varphi, \psi) = (\epsilon \ln(u), -\epsilon \ln(v))$, see [7] for a proof. We can then adapt the previous methods:

- Local perturbation : in the discrete setting, it is impossible to check for the positiveness of the masses of the Laguerre cells since they do not exist. Therefore, we can only choose the weights defined in Proposition 1 (which can also be defined in the discrete setting);
- Rescale : we use it in the same way except that quantities such as centroids and bounding boxes are computed for point clouds;
- Interpolation : in this case, the normalized Lebesgue measure used for the interpolation is supported on a grid (2D or 3D depending on the dimension) that is "around" the source and target point clouds. More precisely, the convex hull of the support of the normalized Lebesgue measure must contain X and Y.

2.5. Pros and cons

To finish this section, we summarize the pros and cons of each method in Table 1. We explain some of the observations below:

- "many iterations" in the Local perturbation method is related to the fact that we do not have explicit bounds on the ϵ parameter. In practice, we choose it to be sufficiently small to reduce the number of iterations in the inner loop;
- "same dimension" in the Rescale method means that X and Y should have the same intrinsic geometry for the method to work i.e. for finding correct values for λ and t;
- precision on the running times of each method will be given in the next section.

Method	Pros	Cons
Local perturbation	fast guarantees SD	quadratic cost many iterations
Rescale	fast guarantees SD	quadratic cost same dimension
Interpolation	any cost guarantees SD	slow

Table 1: Pros and cons of each initialization method. "guarantees X" means that we have theoretical guarantees on the convergence of the method in setting X (where "SD" means semi-discrete and "D" discrete).

3. Numerical results

We now illustrate the three methods we previously described on different examples in 2D and 3D in both the semi-discrete and discrete settings. We also consider some applications.

3.1. Semi-discrete optimal transport

We start by illustrating the convergence speed of the three initialization procedures on different examples in the semi-discrete setting. In all the cases we will consider, the source measure will be supported on a triangulations.

Firstly, we look at two-dimensional examples i.e. when both the source and target measures are supported on 2D domains. In Figure 1, we display the support of the source and target measures as well as the number of Laguerre cells with zero mass during the LOCAL PERTURBATION procedure for a decrement parameter $\epsilon = 10^{-1}$. In Figure 2, we display one setting in which the RESCALE procedure works and one where it doesn't. In the bottom example, the methods to choose t and λ that we proposed are too simple and illustrate the fact that this procedure cannot handle even simple examples (for which the other two methods work). Finally, in Figure 3, we display for the same examples of Figure 1, the results of the INTERPOLATION method. We show the Laguerre cells at the beginning and at the end of the procedure. One can clearly see that, at the end, all the Laguerre cells intersect the support of the source measure.

Secondly, we look at source measures supported on triangulated surfaces. In this setting, the RESCALE method can not be used since, in general, it is not possible to find uniform translation and scaling parameters (t, λ) that moves a point set onto a triangulated surface. That is why we only illustrate the other two methods. In Figure 4, we illustrate the LOCAL PERTURBATION method on two settings. In Figures 5 and 6, we display the results of the INTERPOLATION method. We see that in a few iterations (around 10), we manage to get rid of all the empty Laguerre cells.

We finish by some remarks on the LOCAL PERTURBATION method. We observe that for each case the method converges but the number of iterations remains quite big. This number obviously depends on the decrement parameter ϵ in the algorithm. It should not be chosen too big since it will increase the number of iterations in the inner loop. One can also remark that this method also works when the support of the source measure is disconnected, a case that is not covered by the convergence result presented in [19].

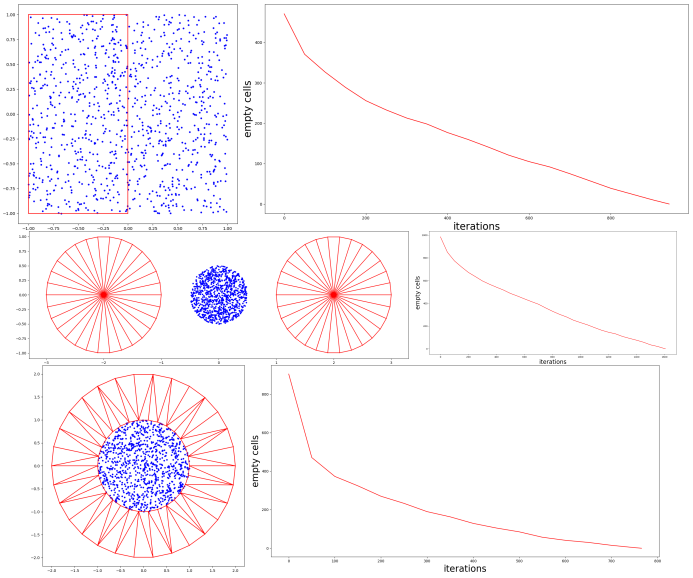

Figure 1: **Left:** setting (the source measure supported on the red domain and the target measure on the blue point set). **Right:** results of the LOCAL PERTURBATION method, we display the number of Laguerre cells with zero mass at each step of the algorithm. Maximal decrement ϵ is set to 10^{-1}.

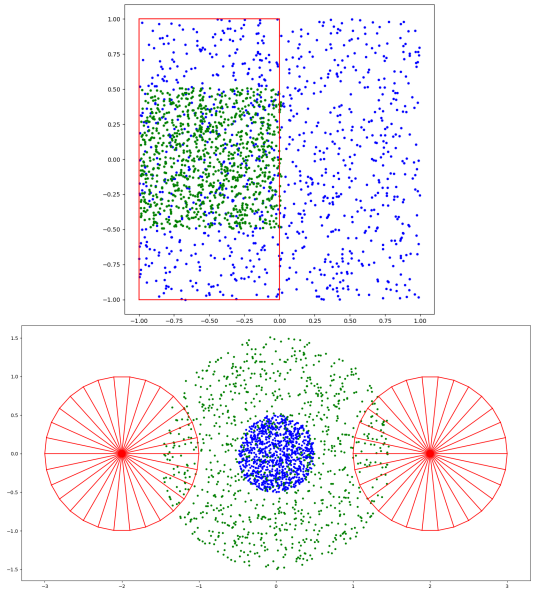

Figure 2: The rescaled target point set is displayed in green in two examples. **Top:** example where the RESCALE method works, **Bottom:** example where it fails (the green set is not contained in the red region).

3.2. Discrete optimal transport

We now look at the case of discrete optimal transport. On Figures 7 and 8, we display the running times of the Sinkhorn-Knopp algorithm (see Algorithm 1) initialized with the weights obtained with the three methods for different regularization parameters and discretizations of measures supported on 2D sets and triangulated surfaces. We observe that the algorithm is faster when using the weights obtained after the RESCALE (red) or the INTERPOLATION (green) method. Choosing the weights from the LOCAL PERTURBATION method always ended in longer running times. One can observe a "bump" in Figure 7 (bottom), we suspect this is due to the specific configuration of randomly sampled points for the corresponding discretization. Furthermore, in Figure 8, only the INTERPOLATION method improves the running time. We think that the LOCAL PERTURBATION method does not improve the running time because of the choice of the initial weights (given by Proposition 1) that make the convergence slower.

3.3. Applications

We now detail some applications where such initialization techniques can be useful.

Rigid matching between a point cloud and a mesh

For the first application, we show how one can use these initialization procedures in some more traditional geometry processing algorithms such as finding the "best" rigid (rotation and translation) transformation between a mesh and a point cloud. To do this, we will use the *OT-ICP* algorithm described in [19] (Section 6.2). In this algorithm, one replaces the Euclidean nearest-neighbor query in the traditional Iterative Closest Point (ICP) algorithm with an optimal transport calculation by associating each point in the point cloud to the centroid of its Laguerre cell on the mesh. The authors show that, despite having a longer running time, the final registration error is smaller. When considering such algorithm, since the points do not lie on the mesh, one is confronted to an initialization problem.

We illustrate the efficiency of the LOCAL PERTURBATION method on different examples, see Figure 9. The point cloud to register

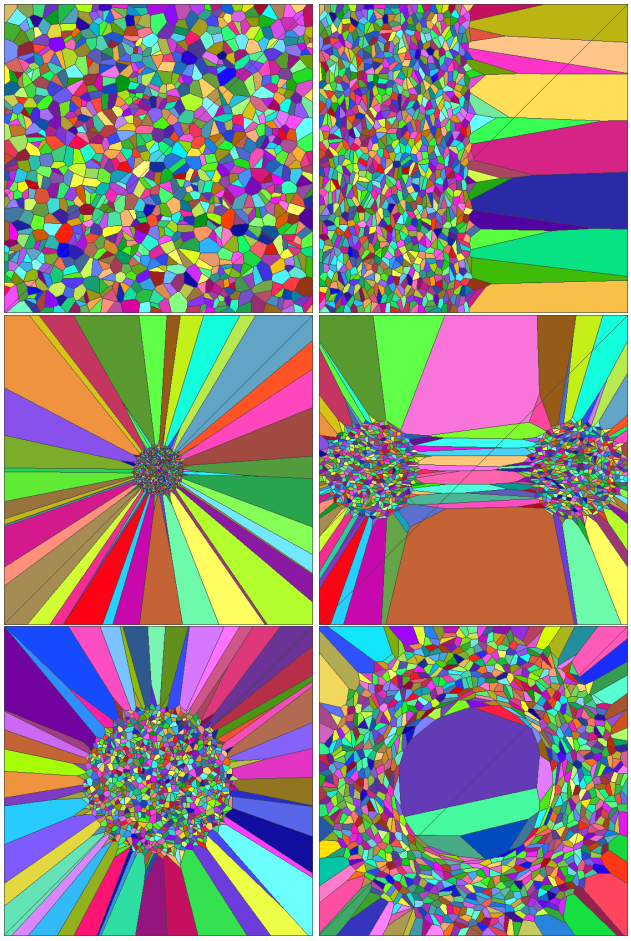

Figure 3: Illustrations of the INTERPOLATION method for the three examples of source and target measures of Figure 1. **Left:** initial Laguerre cells, **Right:** final Laguerre cells. The final Laguerre cells intersect the support of the source measure in the three cases while the initial ones don't.

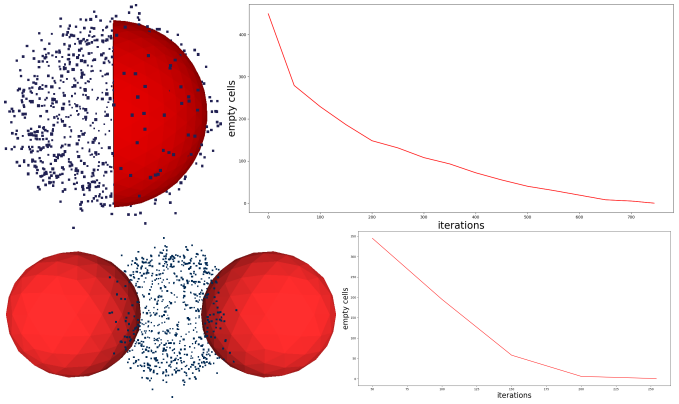

Figure 4: Illustrations of the LOCAL PERTURBATION method for source measures supported on triangulated surfaces. **Left:** setting (the source measure is supported on the red triangulation and the target measure on the blue point cloud). **Right:** evolution of the number of empty Laguerre cells. Maximal decrement ϵ is set to 10^{-1}.

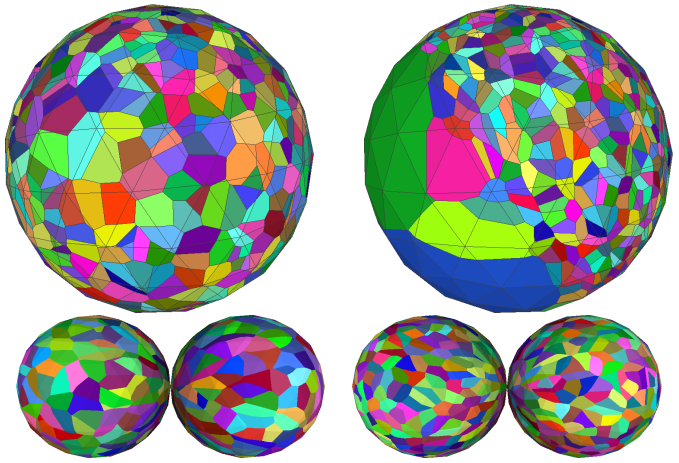

Figure 5: Illustrations of the INTERPOLATION method for source measures supported on triangulated surfaces. **Left:** initial Laguerre cells. **Right:** final Laguerre cells. In the top example, the source measure is supported on the right half of a sphere while in the bottom one, it is supported on the union of two triangulated spheres.

is obtained by sampling points on the target mesh; adding some noise to it; translating it and rotating it by $\pi/2$.

Non-imaging optics

We now detail an application of these initialization procedures in the field of *non-imaging optics*. In this field, one is interested in designing optical components (such as mirrors or lenses) with prescribed reflection or refraction patterns, see Figure 10 for an illustration of one of those problems. Applications of such problems include public lighting where one wants to design optical components in order to decrease light loss and light pollution or for designing car headlights [23] which do not blind incoming cars.

It has been shown [24, 16, 25] that optimal transport allows to solve numerous of these problems in a unified setting. Indeed, the dual variables (the Kantorovich potentials (φ, ψ)) directly gives access to a parametrization of the optical component.

When designing such components, if the size of the light source is too small then it can happen that the Laguerre cells will not intersect it and thus have zero mass, preventing the use of the damped Newton's method. Thus, it is important to have a way to correctly initialize the algorithm. In this context, we used the INTERPOLATION method in a simple example, see Figure 11. In this example, the target light is discrete and is represented as a grid of 32x32 directions, where each direction targets a pixel of the right image in Figure 10. The Laguerre cells are located in the support of the directional light source represented here by a uniform measure on the square $[-1, 0] \times [-1, 1]$. The normalized Lebesgue measure is here supported on $[-1, 1]^2$ where all the directions lie in. We see that at the beginning approximately half of the Laguerre cells do not intersect the support of the light source while after doing the initialization step using the INTERPOLATION method, all Laguerre cells intersect it.

Seismic imaging

Optimal transport has also been shown to be a useful tool in seismic imaging and in particular in the *Full Waveform Inversion* framework. In this framework, one wants to reconstruct the subsurface of the Earth using seismic images. In particular, one needs to be able to compute the error (called *misfit*) between two seismic images. In [26], it has been shown that the distances arising from optimal transport (the so-called *Wasserstein* distances) help in "convexifying" the misfit function and thus avoids local minima. See the top row of Figure 12 for two examples of seismic images. Current algorithms using this misfit function solve optimal transport in the discrete setting using for instance the dual formulation of the 1-Wasserstein distance to handle non-positive measures. The main obstacle in using semi-discrete methods is the fact that

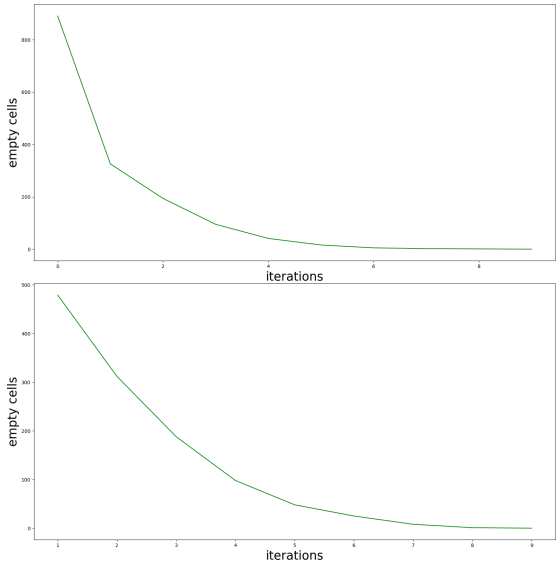

Figure 6: Evolution of the number of empty Laguerre cells for the two examples of Figure 5.

due to the oscillatory nature of the seismic signals, ensuring that all the Laguerre cells have positive mass is not an easy task, see the bottom row of Figure 12.

We show using our initialization procedures that we are able to find such weights and thus compute optimal transport maps. Seismic signals are given as greyscale images. The first thing we do is we convert them into a source and target measures for which we will solve optimal transport. We see the images as height fields (the height representing the amplitude of the signal). We normalize the heights to be in $[-1, 1]$. The target measure is simply the uniform measure supported on this point cloud while the source measure is the uniform measure supported on the Delaunay triangulation of the 2D grid.

Another difficulty raised by the oscillatory natures of the signals is the computation of the Laguerre diagrams becoming expensive, see the benchmarks on the top and middle rows of Figure 13. We use the fact that, for the quadratic cost, one can see the Laguerre diagram as the intersection between a Power diagram and the support of the source measure (here a triangulation). One can then use efficient algorithms to compute this intersection, see [27, 28] for instance. We observe that the computation time is linear in the number of target points and more than linear in terms of the number of source triangles. Nevertheless, we are able to compute initial weights and optimal transport maps, see the bottom row of Figure 13 for one example.

4. Conclusion

In this paper, we presented three different procedures that can be used to improve the performance of numerical methods to solve optimal transport in both the discrete and semi-discrete settings. We also showed numerous numerical examples illustrating the performance of these methods. In the future, we would like to see if we can improve the running time of these methods as well as applying them in other contexts. For instance, we wonder if we can adapt the same kind of techniques developed in [29, 30] for seismic data analysis. in the semi-discrete setting

Acknowledgments. This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir and by the ANR project MAGA (ANR-16-CE40-0014). The author would also like to thank Boris Thibert, Quentin Mérigot for their useful insights and Ludovic Métivier for providing the seismic data.

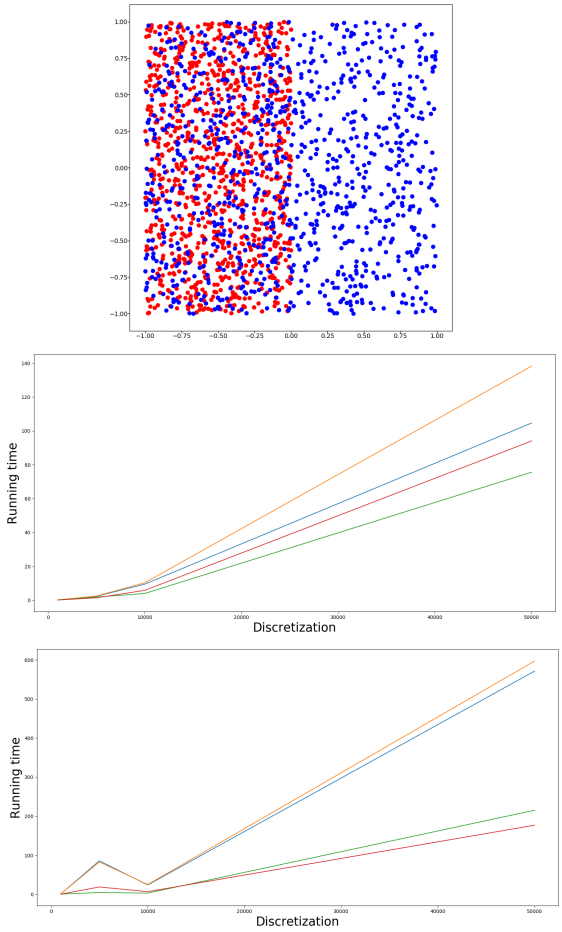

Figure 7: Running times in seconds (middle and bottom) of the Sinkhorn-Knopp algorithm obtained when choosing different initial vectors (u_0, v_0) according to different methods. Color coding of the methods: no initialization = blue; Local perturbation = orange; Rescale = red; Interpolation = green. **Top:** setting (the source measure is sampled in the left part of a square and the target measure on the whole square). **Middle**: $\epsilon = 10^{-1}$. **Bottom**: $\epsilon = 10^{-2}$.

References

[1] G. Tartavel, G. Peyré, Y. Gousseau, Wasserstein loss for image synthesis and restoration, SIAM Journal on Imaging Sciences 9 (4) (2016) 1726–1755.

[2] M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, in: Advances in Neural Information Processing Systems, 2013, pp. 2292–2300.

[3] C. Frogner, C. Zhang, H. Mobahi, M. Araya, T. A. Poggio, Learning with a wasserstein loss, in: Advances in Neural Information Processing Systems, 2015, pp. 2053–2061.

[4] N. Lei, K. Su, L. Cui, S.-T. Yau, X. D. Gu, A geometric view of optimal transportation and generative model, Computer Aided Geometric Design 68 (2019) 1–21.

[5] T. O. Gallouët, Q. Mérigot, A lagrangian scheme à la brenier for the incompressible euler equations, Foundations of Computational Mathematics (2017) 1–31.

[6] C. Villani, Topics in optimal transportation, Vol. 58, American Mathematical Soc., 2003.

[7] F. Santambrogio, Optimal transport for applied mathematicians, Birkäuser, NY.

[8] B. Lévy, E. L. Schwindt, Notions of optimal transport theory and how to implement them on a computer, Computers & Graphics 72 (2018) 135–148.

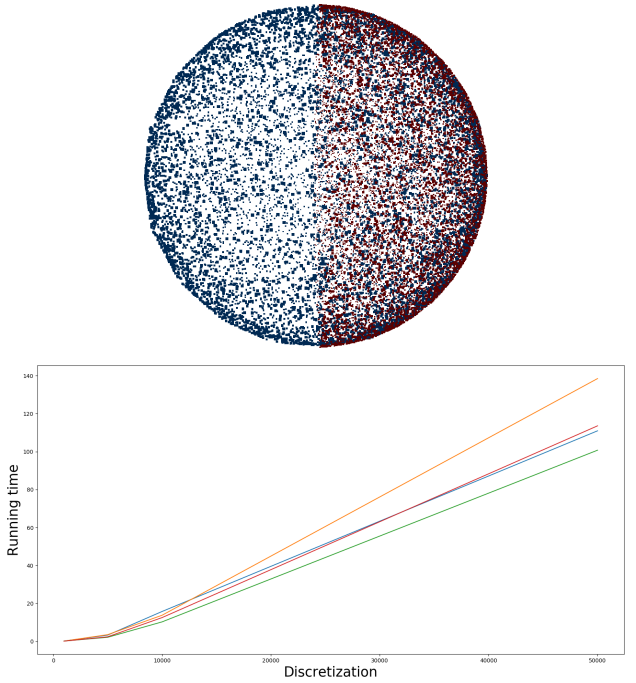

Figure 8: Running times in seconds of the Sinkhorn-Knopp algorithm obtained when choosing different initial weights according to different methods. Color coding of the methods : no initialization = blue; LOCAL PERTURBATION = orange; RESCALE = red; INTERPOLATION = green. **Top:** setting (the source measure is sampled in the right hemisphere and the target measure on the whole sphere). **Bottom**: $\epsilon = 10^{-1}$.

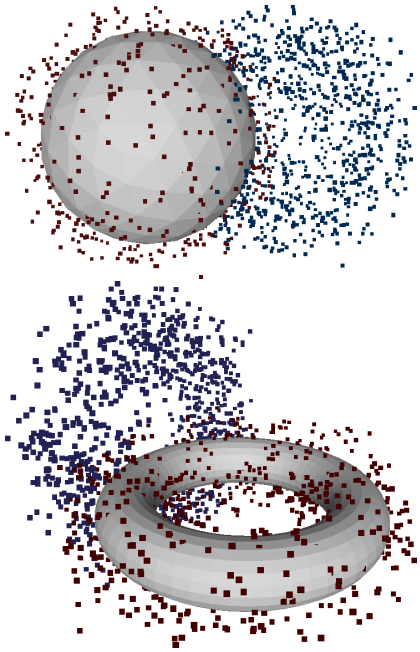

Figure 9: Settings in the OT-ICP algorithm: the target mesh is in gray while the point cloud to be registered is in blue (1000 points). The red point cloud is the final registered point cloud. The maximal error for the optimal transport problem is set to $\eta = 10^{-6}$.

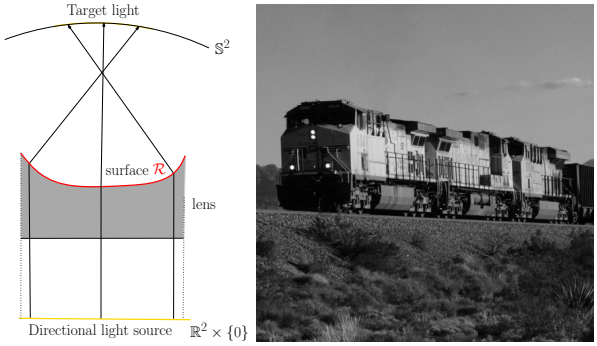

Figure 10: **Left:** Example of optical component design problem. Here, one is interested in designing the outer surface $\mathcal{R}$ of a lens that refracts the input directional light source (that emits parallel rays) onto a prescribed target light at infinity (i.e. supported on the 3D sphere $\mathbb{S}^2$). The refraction index of the lens is an input. **Right:** example of target image. Each pixel encodes a direction whose pixel intensity is the value of the measure in this direction.

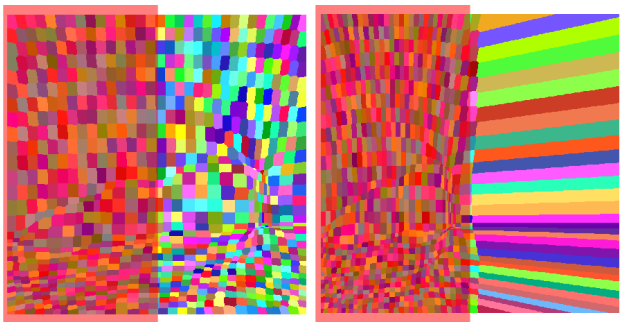

Figure 11: INTERPOLATION method in an optical component design problem. We display the initial and final Laguerre cells. The directional light source is supported on the left part of the square (the red overlay). Interpolating between this measure and the Lebesgue measure on the whole square $[-1, 1]^2$ gives us Laguerre cells intersecting the support of the light.

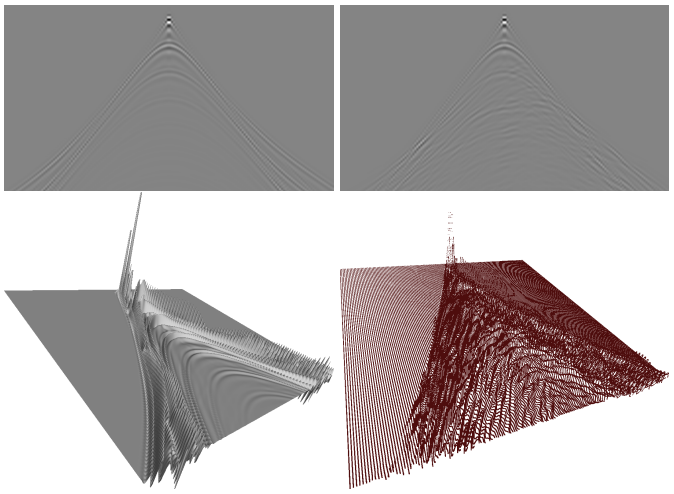

Figure 12: **Top:** Two examples of seismic images for 2000 receptors and 168 timesteps. **Bottom:** Seismic signals converted to a source (triangulated surface) and target (discrete) measures for semi-discrete optimal transport.

[9] K. Su, W. Chen, N. Lei, L. Cui, J. Jiang, X. D. Gu, Measure controllable volumetric mesh parameterization, Computer-Aided Design 78 (2016) 188–198.

[10] G. Monge, Mémoire sur la théorie des déblais et des remblais, Histoire de l'Académie Royale des Sciences de Paris.

[11] L. Kantorovitch, On the translocation of masses, Management Science 5 (1) (1958) 1–4.

[12] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, L. Guibas, Convolutional wasserstein distances: Efficient optimal transportation on geometric domains, ACM Transactions on Graphics (TOG) 34 (4) (2015) 66.

[13] J. Feydy, B. Charlier, F.-X. Vialard, G. Peyré, Optimal transport for dif-

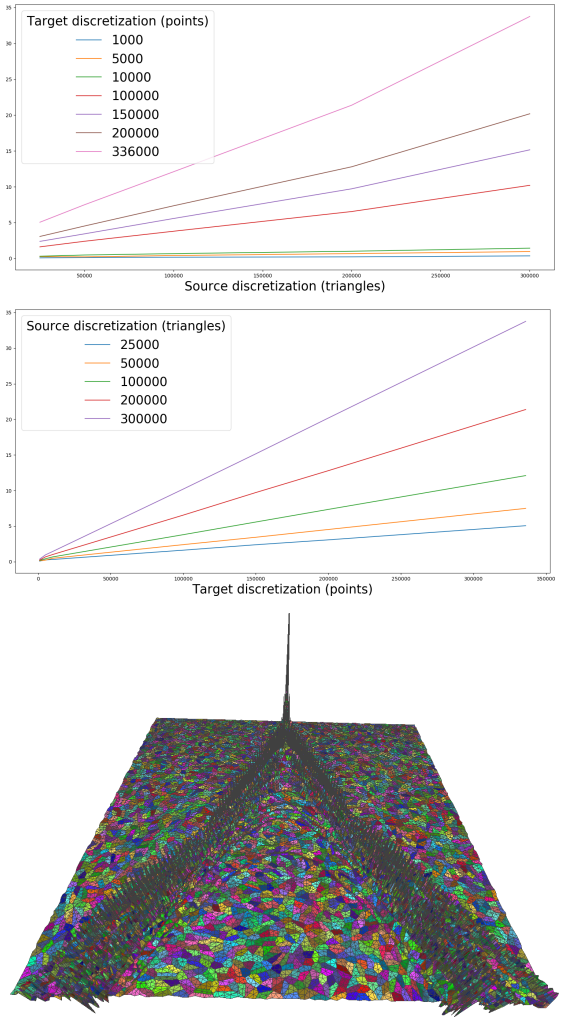

Figure 13: **Top and middle:** Running times in seconds for the computation of one Laguerre diagram with various discretizations of the source and target measures (number of triangles for the source measure and number of points for the target one). Top: fixed target discretization, increasing source discretization. Middle: fixed source discretization, increasing target discretization. **Bottom:** Optimal transport map represented as the collection of Laguerre cells on the support of the source measure.

feomorphic registration, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 291–299.

[14] F. Aurenhammer, F. Hoffmann, B. Aronov, Minkowski-type theorems and least-squares clustering, Algorithmica 20 (1) (1998) 61–76.

[15] F. de Goes, K. Breeden, V. Ostromoukhov, M. Desbrun, Blue noise through optimal transport, ACM Transactions on Graphics 31 (6) (2012) 171.

[16] F. de Goes, C. Wallez, J. Huang, D. Pavlov, M. Desbrun, Power particles: an incompressible fluid solver based on power diagrams., ACM Trans. Graph. 34 (4) (2015) 50–1.

[17] J.-M. Mirebeau, Discretization of the 3D Monge-Ampere operator, between Wide Stencils and Power Diagrams, arXiv preprint arXiv:1503.00947.

[18] J. Kitagawa, Q. Mérigot, B. Thibert, A Newton algorithm for semi-discrete optimal transport, arXiv preprint arXiv:1603.05579.

[19] Q. Mérigot, J. Meyron, B. Thibert, An algorithm for optimal transport between a simplex soup and a point cloud, SIAM Journal on Imaging Sciences 11 (2) (2018) 1363–1389.

[20] X. Gu, F. Luo, J. Sun, S.-T. Yau, Variational principles for minkowski type problems, discrete optimal transport, and discrete monge–ampère equations, Asian Journal of Mathematics 20 (2) (2016) 383–398.

[21] Q. Mérigot, A multiscale approach to optimal transport, in: Computer Graphics Forum, Vol. 30, Wiley Online Library, 2011, pp. 1583–1592.

[22] P. M. M. de Castro, Q. Mérigot, B. Thibert, Far-field reflector problem and intersection of paraboloids, Numerische Mathematik (2015) 1–23.

[23] J. André, D. Attali, Q. Mérigot, B. Thibert, Far-field reflector problem under design constraints, International Journal of Computational Geometry & Applications 25 (02) (2015) 143–162.

[24] X.-J. Wang, On the design of a reflector antenna, Inverse problems 12 (3) (1996) 351.

[25] J. Meyron, Q. Mérigot, B. Thibert, Light in power: a general and parameter-free algorithm for caustic design, in: SIGGRAPH Asia 2018 Technical Papers, ACM, 2018, p. 224.

[26] L. Métivier, R. Brossier, Q. Mérigot, E. Oudet, J. Virieux, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society 205 (1) (2016) 345–377.

[27] B. Lévy, A numerical algorithm for L^2 semi-discrete optimal transport in 3D, ESAIM: Mathematical Modelling and Numerical Analysis 49 (6) (2015) 1693–1715.

[28] M. Sainlot, V. Nivoliers, D. Attali, Restricting voronoi diagrams to meshes using corner validation, in: Computer Graphics Forum, Vol. 36, Wiley Online Library, 2017, pp. 81–91.

[29] L. Métivier, A. Allain, R. Brossier, Q. Mérigot, E. Oudet, J. Virieux, A graph-space approach to optimal transport for full waveform inversion, in: SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 2018, pp. 1158–1162.

[30] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, D. Slepčev, A transportation L^p distance for signal analysis, Journal of Mathematical Imaging and Vision 59 (2) (2017) 187–210.