DISCRETE OPTIMAL TRANSPORT: COMPLEXITY, GEOMETRY AND APPLICATIONS

QUENTIN MÉRIGOT AND ÉDOUARD OUDET

Abstract. In this article, we introduce a new algorithm for solving discrete optimal transport based on iterative resolutions of local versions of the dual linear program. We show a quantitative link between the complexity of this algorithm and the geometry of the underlying measures in the quadratic Euclidean case. This discrete method is then applied to investigate to wo optimal transport problems with geometric flavor : the regularity of optimal transport plan on oblate ellipsoids, and Alexandrov's problem of reconstructing a convex set from its Gaussian measure.

1. Introduction

Numerical methods for optimal transport are currently restricted to very specific settings: the measures involved should have a continuous density, the space should be a flat torus or an Euclidean space and the cost function should be given by the square of the Euclidean distance (see below for a survey of these methods and their exact application domain). In many applications of optimal transport in geometry, however, the underlying cost function is non-Euclidean. This happens for instance when the underlying space is a more general Riemannian manifold and the cost function is the squared Riemmanian distance. More general cost functions also arise in the reformulation of some classical inverse problems arising in convex geometry into the language of optimal transport. For instance, the cost function that naturally appears when reformulating Alexandrov's problem [23, 5] is $c(u,v) = -\log(\max(\langle u|v\rangle, 0))$, and the underlying space is the unit d-sphere. Similar cost functions also appears in inverse problems coming from geometric optics [11]. In addition to non-Euclidean cost functions, it is sometimes natural in these geometric situations to consider optimal transport between measures that are not absolutely continuous with respect to the Lebesgue measure on the underlying space. Examples of such measures include the curvature measures of singular convex sets. These difficulties make it impossible to use currently existing numerical approaches for optimal transport in many geometric problems.

This article makes two main contributions. First, we propose a simple algorithm based on local linearizations in order to solve discrete optimal transport. Our main contribution is an upper bound on the complexity of each local linearizations in term of the geometry of the underlying measures. This bound holds when the cost function is the squared Euclidean norm, and is obtained as the consequence of a theorem from convex analysis due to Bronsted and Rockafellar. This gives a first quantitative explanation

of the (empirically recorded) hardness of solving discrete optimal transport problems involving highly concentrated measures. The second contribution consists in numerical experiments that illustrate the effectiveness of simple discretizations of optimal transport in order to solve continuous geometric questions.

1.1. **Overview of numerical methods for solving optimal transport.**

Continuous methods. There exist many PDE-inspired numerical methods to solve the quadratic transport problem. The first to have been proposed is an augmented Lagrangian method of Benamou and Brenier [3] that solves the time-continuous optimal transport problem on the plane, also recovering the optimal "static" transport map. Loeper and Rapetti[21] proposed to solve optimal transport between the uniform measure and a measure with positive density on the torus by a direct resolution of the Monge-Ampère equation corresponding to mass preservation under periodic boundary conditions. Very recently Benamou, Froese and Oberman [4] propose a method to solve the Monge-Ampère equation under the boundary conditions that arise in the (non-periodic) Euclidean optimal transport problem.

Semi-discrete optimal transport. Instead of solving the optimal transport problem between absolutely continuous measures, one can approximate the target measure by a sum of Dirac masses. This approach has been used first as a theoretical tool by proposed by Pogorelov [25], and many numerical implementations have been proposed [12, 24, 2, 22]. In order to apply this approach, one needs to compute generalized Laguerre diagrams in an efficient way, e.g. almost linear in the number of points. This practically limits this approach to 2D problems, and for the squared Euclidean distance.

Discrete optimal transport. The simplest discretization of the optimal transport problems is known in the combinatorial optimization community as the *linear assignement problem.* Given an integer-valued square matrix $(c_{ij})_{1\leqslant i,j\leqslant N}$, the question is to find a permutation σ of the set $\{1,\dots,N\}$ that minimizes the total transport cost $\sum_{1\leqslant i\leqslant N} c_{i\sigma(i)}$. The linear assignment problem is a standard problem in combinatorial optimization and many algorithms have been developed to solve it [9], among which the so-called Hungarian method or the auction algorithm [6] and their variants. The best worst-case complexity for solving the linear assignment problem for a *dense* $N\times N$ integer-valued cost matrix is of order $\mathrm{O}(N^{5/2}\log(NC))$, where C is the value of the maximum entry in the cost matrix. Note that this worst-case complexity is only an upper-bound on the running time of the best algorithm currently known, and there is no known lower bound on the complexity of the problem but the trivial ones. This worst-case bound is too general to fully exploit the geometry of the problem and is therefore very often not representative of the running time of the algorithm.

Some algorithms [**?**] exploit the geometry of the cost function to improve these upper bounds. However, there seems to be no complexity analysis that exploits the "geometry" of the two involved point sets, such as their concentration on lower dimensional subsets, even though this geometry has a *huge* impact on the actual running times, up to many order of magnitude.

1.2. **Contributions.** We introduce in an iterative method made of local linearizations in order to solve the discrete optimal transportation problem. In the linear assignment case, this means solving the linear program that appear, which has $2N$ variables and N^2 constraints, to a sequence of linear programs with fewer constraints. Our main theoretical result is an upper bound on the complexity of these local linearization in the squared Euclidean case, as measured by the number of constraints. It turns out that in this case, the number of contraints that appear is equal to the cardinality of the *weighted ε-neighborhood graph*. Given for two point sets X and Y in $\mathbb{R}^d$, and a weight function $w : Y \to \mathbb{R}$, this graph is defined as follows:

$$G_\varepsilon(w) = \left\{ (x,y) \in X \times Y;\ \|x-y\|^2 + w(y) \leqslant \min_{z\in Y}[\|x-z\|^2 + w(z)] + \varepsilon \right\}.$$

We show that the maximum cardinality of $G_\varepsilon(w)$ can be linear in the cardinal $N = |X| = |Y|$ when these sets satisfy some simple geometric assumptions. This result (Theorem 4.1) is obtained as a consequence of a theorem on the ε-differential of convex functions due to Bronsted and Rockafellar. To the best of our knowledge, this is the first quantitative highlight of a link between the complexity of an algorithm designed to solve discrete optimal transport problems and the geometry of the underlying measures.

In the last section we illustrate how very simple discretizations can yield approximate solutions to two continuous optimal transport problems with geometric flavor, that are currently out of reach for numerical PDE methods. In these applications, the probability measures that are involved are replaced by uniform probability measures on discrete point sets, and we use our algorithm to solve the discrete optimal transport problems.

- In a first illustration we show how our discrete approach can give qualitative information when studying the regularity of optimal transports. More precisely, as it has been proved in [7, 13], we illustrate numerically the fact that discontinuous transport plans can occur between non-vanishing densities even on simple convex manifolds like oblate ellipsoids.
- Our second example is related to Alexandrov's problem, i.e. the reconstruction a convex surface from its Gaussian curvature measure. The reconstructed surface has been characterized in [23, 5] in term of optimal transport potentials, and we show how our method can be used to practically solve this inverse problem. We illustrate this reconstruction procedure on the very degenerated case (in the sense of the concentration of the measures) of the icosahedron and on the reconstruction of a convex body with singular edges.

2. Background on optimal transport

2.1. **Discrete optimal transport and assignment problem.** Suppose we are given two probability measures μ and ν supported on two finite sets X, Y, and can be written $\mu = \sum_{x\in X} \mu_x \delta_x$ and $\nu = \sum_{y\in Y} \nu_y \delta_y$ for some non-negative set of masses (μ_x) and (ν_y). In this setting, a *transport plan* between μ and ν is described by a matrix $(\pi_{xy})_{(x,y)\in X\times Y}$ with non-negative

entries that satisfies the two following set of constaints

$$\forall x \in X, \sum_{y \in Y} \pi_{xy} = \mu_x \text{ and } \forall y \in Y, \sum_{x \in X} \pi_{xy} = \nu_y.$$

The set of transport plans between μ and ν is denoted by $\Gamma(\mu, \nu)$.

We are given a *cost function* $\mathrm{c} : X \times Y \to \mathbb{R}$ on the product space, i.e. the value $c(x, y)$ measures how costly it is to move mass from x to y. The *optimal transport problem* between μ and ν is the following optimization problem

$$\min \left\{ \sum_{x \in X} \mathrm{c}(x, y)\pi_{xy};\ \pi \in \Gamma(\mu, \nu) \right\}. \tag{P1}$$

Note that the set of transport plans is a polyhedron and the optimized function linear thus implying that the optimal transport problem is a linear programming problem. The following theorem of Kantorovich characterizes the dual of this linear program.

THEOREM 2.1 (Kantorovich duality). *The dual of the linear programming problem* (P1) *is given by*

$$\begin{aligned} &\max_{v,w} \sum_{x \in X} v(x)\mu_x - \sum_{y \in Y} w(y)\nu_y \\ &\textit{where } v : X \to \mathbb{R}, w : Y \to \mathbb{R} \textit{ are s.t.} \\ &\qquad \forall (x, y) \in X \times Y,\ v(x) - w(y) \leqslant \mathrm{c}(x, y). \end{aligned} \tag{P2}$$

Moreover, this linear program is equivalent to the following unconstrained concave maximization problem $\max_w \Phi(w)$ *where*

$$\Phi(w) := \sum_{x \in X} \left[\min_{y \in Y} w(y) + \mathrm{c}(x, y) \right] \mu_x - \sum_{y \in Y} w_y \nu_y, \tag{P3}$$

and the supremum is taken on any weight function $w : Y \to \mathbb{R}$.

Linear assignment problem. When X and Y have the same cardinality N and the two measures μ and ν are uniform on the sets X and Y, the optimal transport problem is equivalent to the *(linear) assignment problem*, which is well-known in combinatorial optimization. It consists in finding a bijection $\sigma : X \to Y$ which realizes the minimum

$$\min_{\sigma} \sum_{x \in X} \mathrm{c}(x, \sigma(x)) \text{ where } \sigma \in \text{ is a bijection between } X \text{ and } Y. \tag{PA}$$

While the theoretical results below apply to the general transport problem as well as to the linear assignment problem, in the applications of Section 5 we only make use our algorithm in the special case of the linear assignment problem. We make no claim on the efficiency of the local linearization algorithm presented below for the general discrete optimal transport problem.

3. Linear assignment through local linearizations

Given a weight vector w and a positive number ε, we describe in this section a simple linear program whose solution allows to recover the maximum of the concave function Φ defined in (P3) over the hypercube $w + [0, \varepsilon]^N$. In

order to define this linear program, we consider the following subset of the product set $X \times Y$

$$\mathrm{G}_\varepsilon(w) = \{(x,y) \in X \times Y; w(y) + c(x,y) \leqslant \min_{z \in Y} w(z) + c(x,z) + \varepsilon\}.$$

For instance, when c is the squared Euclidean distance on $\mathbb{R}^d$ and w vanishes, $\mathrm{G}_\varepsilon(w)$ coincides to the more usual ε-nearest neighbor graph. Following this terminology, one could call the generalization to non-vanishing weights the weighted ε-nearest neighbor graph.

3.1. **Definition of the local linearization.** The *local linearization* of problem (P2) around a weight vector w and at a certain scale $\varepsilon > 0$ is defined by the following linear programming problem:

$$\max_{v,\delta} \sum_{x \in X} v(x)\mu_x - \sum_{y \in Y} (w(y) + \delta(y))\nu_y \tag{P2$^\varepsilon_w$}$$

where $v : X \to \mathbb{R}$ and $\delta : Y \to \mathbb{R}$ are functions satisfying the two following sets of contraints

$$\begin{aligned} &\forall (x,y) \in \mathrm{G}_\varepsilon(w), && v(x) - (w(y) + \delta(y)) \leqslant \mathrm{c}(x,y) \\ &\forall y \in Y, && 0 \leqslant \delta(y) \leqslant \varepsilon \end{aligned}$$

Note that this local linearization tries to make a trade-off between the number of constraints, which grows to $|X| \cdot |Y|$ as ε grows, and the size of the hypercube $[0,\varepsilon]^{|X|}$ to which the offset weight vector vector δ belongs.

The following easy lemma shows that (P2$^\varepsilon_w$) is indeed a local linearization of (P2)

LEMMA 3.1. *Any pair of vectors $(v, w+\delta)$ which satisfies the constraints of the local linearization* (P2$^\varepsilon_w$) *also satisfy the constraints of* (P2).

Proof. Fix a point x in X, and order the points in Y by increasing values of be an ordering $w(y) + c(x,y)$, i.e.

$$w(y_1) + c(x,y_1) \leqslant w(y_2) + c(x,y_2) \leqslant \dots$$

Now, consider now a point y such that (x,y) does not belong to $\mathrm{G}_\varepsilon(w)$, i.e. such that

$$w(y) + c(x,y) > \varphi_{Y,w}(x) + \varepsilon = c(x,y_1) + w(y_1) + \varepsilon.$$

Using this inequality, we have:

$$v(x) - (w(y) + \delta(y)) < v(x) + c(x,y) - (c(x,y_1) + w(y_1) + \varepsilon + \delta(y))$$

Since the pair of vectors $(v, w+\delta)$ satisfies the constraints of (P2$^\varepsilon_w$), we know that $v(x) - w(y_1) - c(x,y_1) \leqslant \delta(y_1)$ and

$$v(x) - (w(y) + \delta(y)) \leqslant c(x,y) + \delta(y_1) - \delta(y) - \varepsilon.$$

Using the fact that $\delta(y_1)$ and $\delta(y)$ belong to $[0,\varepsilon]$, we see that the pair $(v, w+\delta)$ also satisfies all the constraints of (P2). □

The following Corollary characterizes the cases where the maximum of (P2$^\varepsilon_w$) yields a global maximizer of (P2), thus providing a stopping criterion for our algorithm.

COROLLARY 3.2. *Suppose the maximum of* $(\mathrm{P2}^{\varepsilon}_{w})$ *is attained at an admissible pair* $(v, w+\delta)$ *such that for every point* y *in* Y, $\delta(y)$ *belongs to* $(0,\varepsilon)$. *Then, the weight vector* w *is a global maximizer of* (P3).

Proof. By the previous lemma, any maximizer $(v, w+\delta)$ of $(\mathrm{P2}^{\varepsilon}_{w})$ is admissible for the problem (P3). If for every point y the value of $\delta(y)$ belongs to $(0,\varepsilon)$, this pair satisfies the same optimality conditions as optimal vectors of Problem (P3). Thus, w it is a global maximizer of (P3) by concavity. □

Remark 3.1. In [10], Buš and Tvrdík describe a heuristic to increase the speed of the Berstekas' auction algorithm, which exploits the fact that the weight vectors do not change much between two successive steps of the auction algorithm. This heuristic is quite close to the local linearisation proposed here. It is therefore likely that the theoritecal results below on the size of $\mathrm{G}_{\varepsilon}(w)$ in the squared distance case could be used to explain the practical efficiency of this heuristic.

3.2. **Algorithm.** This local linearization can be applied iteratively in order to construct a solution of the optimal transport problem (Algorithm 1). The following Lemma asserts that the algorithm terminates in a finite number of steps.

Algorithm 1 Discrete optimal transport through local linearizations.

Input: A cost matrix (C_{xy}), a parameter $\varepsilon > 0$.
Initialization: $w_0 \leftarrow 0$
Repeat: $w_{i+1} \leftarrow w_i + \delta_i$, where δ_i is a solution of $(\mathrm{P2}^{\varepsilon}_{w_i})$
  until δ_i meets the stopping criterion of Corollary 3.2.
Output: An optimal weight vector $w := w_i$.

LEMMA 3.3. *For any fixed positive* ε, *and any starting point* w_0, *the sequence* (w_i) *defined as above converges to a global maximizer of* (P3) *in a finite number of steps.*

Proof. Thanks to Corollary 3.2, we know that if $w_{i+1} = w_i + \delta_i$ is not the global maximum, then δ_i belongs to the boundary of the box $[0,\varepsilon]^M$, and there exists no constant c such that $w + c$ belongs to the interior $(0,\varepsilon)^M$. This implies that there exists two points y, y' in Y such that $\delta_i(y) = \varepsilon$ and $\delta_i(y') = 0$, thus ensuring that the distance between w_i and w_{i+1} is uniformly bounded from below by $\varepsilon/2$. One concludes by using the fact that the graph of the function optimized in (P3) is made of a finite number of linear pieces, implying a (uniform) minimum increase in $\Phi(w_{i+1}) - \Phi(w_i)$ at each step before convergence, where Φ is the function defined in (P3). □

3.3. **Implementation details.** Our implementation follows the approach described in the previous paragraph, with two important differences in the choice of the initial weight vector, and in the practical construction of the local linearization.

Initialization through convex optimization. The dual form of our optimization problem can be seen as a large scale non-smooth concave maximization problem. A surprisingly efficient method for minimizing non-smooth functions is the use of quasi Newton methods. Indeed, different authors pointed out recently [15, 18, 17] that variable metric algorithms may produce in some cases sequences which converge to an optimal point in the sense of Clarke. The mathematical analysis of this good behavior has just been initiated in recent papers of Overton [18, 17].

A detailed study of the application of quasi-Newton methods to optimal transportation would be out of the scope of that paper. However, we point out that L-BFGS (low memory version of Broyden-Fletcher-Goldfarb-Shanno algorithm), in the context of optimal transportation, gives a rather efficient way to obtain a good starting weight vector for the local linearizations approach.

Construction of the local linearization. Instead of relying on an ε parameter, we use a "budget" on the number of linear constraints to be used in the linear program. More precisely, at every local linearization step, we start with a weight vector $w : Y \to \mathbb{R}$. We let M be the number of points in Y. For every point x in X, we compute a map $\sigma_x : \{1, \dots, M\} \to Y$ ordering the points of Y by increasing values of $c(x, y) + w(y)$, i.e.

$$c(x, \sigma_x(1)) + w(\sigma_x(1)) \leqslant \dots \leqslant c(x, \sigma_x(M)) + w(\sigma_x(M)). \tag{3.1}$$

In the full linear program (P2), each couple (x, y) in $X \times Y$ yields a linear constraint on the dual variables. For the purpose of our local linearisation, we choose the number of constraint I_x for each point x so as to maximize:

$$\varepsilon := \min_{x \in X} (c(x, \sigma_x(I_x)) - c(x, \sigma_x(1))) \tag{3.2}$$

under the budget constraint $\sum_{x \in X} I_x \leqslant B$, where B is the maximum number of constraints in the linear program. This is efficiently done using a priority queue. In practice setting $B = |X|\,|Y|^{1/2}$ offers a good trade-off between the cost of solving the number of linear constraints versus maximizing the value of ε. Note that if ε vanishes, one would have to increase the value of B – however, we never encountered that case in practice.

4. Complexity of a local linearization for the quadratic cost

In this paragraph, we study the complexity of the local linearization, as measured by the number of constraints in the linear program ($\mathrm{P2}^{\varepsilon}_{w}$). Said otherwise, we want to upper bound the cardinality of the set $\mathrm{G}_{\varepsilon}(w)$. The theorem below concerns the case where the two point sets X, Y are finite subsets of the Euclidean space $\mathbb{R}^d$, and when the transport cost is given by the squared Euclidean distance $c(x, y) := \frac{1}{2} \|x - y\|^2$.

Before stating the theorem, let us present two examples of point distributions that can cause the set of constraints $\mathrm{G}_{\varepsilon}(w)$ to be equal to the whole product $X \times Y$, even when w vanishes:

(i) A very simple case is when the target point cloud is highly concentrated. Consider $X \subseteq \mathbb{R}^d$, y_0 a point and r the maximum distance

between X and y_0. Consider $\varepsilon < r$ and let Y be a point cloud contained in the ball $\mathrm{B}(y_0, \frac{\varepsilon}{3})$. Then, for a vanishing weight vector w, one has $G_\varepsilon(w) = X \times Y$.

(ii) The second, less trivial example, occurs when X and Y belong to two orthogonal affine subspaces. For instance, let X, Y be two subsets of $\mathbb{R}^3$ such that (i) X is contained in the first coordinate axis $(0x)$ and (ii) Y is contained in the circle obtained by intersecting the unit sphere with the yz-plane. Then, even for ε equal to 0 and a vanishing weight vector, one has $G_\varepsilon(w) = X \times Y$.

These two phenomena are captured by the constants that appear in the following upper bound on the cardinal of $\mathrm{G}_\varepsilon(w)$. The term $(mn)^{3/2}$ measures the concentration of the point sets X and Y. The term involving the Minkowski sum of Ω and Ω', defined by $\Omega + \Omega' = \{x + y; (x, y) \in \Omega \times \Omega'\}$, measures the "approximate orthogonality" of the two supports.

THEOREM 4.1. *Let Ω and Ω' be two open subsets of $\mathbb{R}^d$, and $X \subseteq \Omega$, $Y \subseteq \Omega'$ two sets of cardinal N. Assume moreover that X and Y are evenly distributed in Ω and Ω' in the following sense:*

$$\forall x \in \Omega,\ 1 \leqslant \left|\mathrm{B}(x, \sqrt{\varepsilon}) \cap X\right| \leqslant n$$
$$\forall y \in \Omega',\ 1 \leqslant \left|\mathrm{B}(y, \sqrt{\varepsilon}) \cap Y\right| \leqslant m$$

Then, for any weight vector $w : Y \to \mathbb{R}$, we have the following upper bound on the number of constraints appearing in the local linearization:

$$|\mathrm{G}_\varepsilon(w)| \leqslant \mathrm{const}(d) \left(\frac{\mathcal{H}^d(\Omega + \Omega')}{(\mathcal{H}^d(\Omega)\mathcal{H}^d(\Omega))^{1/2}} \right) (nm)^{3/2} N.$$

The proof of this theorem is postponed to the end of the section.

4.1. **Relation to ε-subdifferentials.** We start by two Lemmas, which constitute the continuous version of Theorem 4.1. Recall that the *ε-subdifferential of a convex function $\psi : \mathbb{R}^d \to \mathbb{R}$* can be defined in the following way [8]:

$$\partial^\varepsilon \psi(x) := \{y \in \mathbb{R}^d;\ \forall h \in \mathbb{R}^d,\ \langle x|y\rangle + \psi(x) \leqslant \psi(x + h) + \varepsilon\}$$

When ε vanishes, this set is called the subdifferential of ψ at x and is simply denoted by $\partial\psi(x)$. The following Lemma makes explicit the relation between the set $G_\varepsilon(w)$ introduced for the local linearization and the graph of the ε-subdifferential.

LEMMA 4.2. *Given a weight vector $w : Y \to \mathbb{R}$, consider the function*

$$\psi : x \mapsto \frac{1}{2}\|x\|^2 - \min_{y \in Y} \left(\frac{1}{2}\|x - y\|^2 + w(y) \right).$$

Then the function ψ is convex and we have the following inclusion

$$\mathrm{G}_\varepsilon(w) \subseteq \mathrm{Graph}(\partial^\varepsilon \psi) := \{(x, y) \in X \times Y;\ y \in \partial^\varepsilon \psi(x)\}.$$

Proof. The convexity of the function ψ is easy to check. By definition a couple (x, y) of belongs to $\mathrm{G}_\varepsilon(w)$ if

$$w(y) + \frac{1}{2}\|x - y\|^2 \leqslant \min_{y \in Y} \frac{1}{2}\|x - y\|^2 + w(y) + \varepsilon,$$

thus implying

$$\begin{aligned}\psi(x+h) &= \frac{1}{2}\|x+h\|^2 - \varphi_{Y,w}(x+h)\\ &\geqslant \frac{1}{2}\|x+h\|^2 - \left(\frac{1}{2}\|x+h-y\|^2 + w(y)\right)\\ &\geqslant \frac{1}{2}\|x\|^2 - \varphi_{Y,w}(x) - \varepsilon + \langle x|h\rangle = \psi(x) - \varepsilon + \langle x|h\rangle.\end{aligned}$$

This shows the desired inclusion. □

The proof of the second Lemma relies on the following theorem of Bronsted and Rockafellar [8]: *Consider a convex function $\varphi : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ and a couple (x, y) such that y belongs to $\partial^\varepsilon\varphi(x)$. Then, for every positive ε, there exists $(x_\varepsilon, y_\varepsilon)$ with $y_\varepsilon \in \partial\varphi(x_\varepsilon)$ and such that $\|x - x_\varepsilon\| \leqslant \sqrt{\varepsilon}$ and $\|y_\varepsilon - y\| \leqslant \sqrt{\varepsilon}$.*

LEMMA 4.3. *Let $\psi : \Omega \to \mathbb{R}$ be a convex function such that the subdifferential of any point x in Ω is contained in Ω'. Then, for any positive number ε,*

$$\mathcal{H}^d \otimes \mathcal{H}^d(\mathrm{Graph}(\partial^\varepsilon\psi)) \leqslant \mathrm{const}(d)\mathcal{H}^d(\Omega + \Omega')\varepsilon^{d/2}.$$

Proof. Consider the graph of the ε-subdifferential $H_\varepsilon := \mathrm{Graph}(\partial^\varepsilon\psi)$, the graph of the subdifferential $H = H_0$, and the tubular neighborhood

$$H^\varepsilon := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d; \exists (x', y') \in H \text{ s. t. } \left\|(x, y) - (x', y')\right\| \leqslant \varepsilon\}.$$

The theorem of Bronsted and Rockafellar asserts that H_ε is included in the tubular neighborhood $H^{\sqrt{2\varepsilon}}$, and our goal is then to bound the volume of this latter set.

Consider the map $\Sigma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ defined by $\Sigma(x, y) = x + y$. For any two pair of points (x, y) and (x', y') in H, the convexity of ψ implies that $\langle x - x'|y - y'\rangle \geqslant 0$. Consequently,

$$\begin{aligned}\left\|\Sigma(x, y) - \Sigma(x', y')\right\|^2 &= \left\|x - x'\right\|^2 + \left\|y - y'\right\|^2 + 2\langle x - x'|y - y'\rangle\\ &\geqslant \left\|(x, y) - (x', y')\right\|^2. \qquad (4.3)\end{aligned}$$

In other words, the restriction of the map Σ on the graph H is one-to-one and its reciprocal is 1-Lipschitz. Denoting by $\mathcal{N}(X, \eta)$ the minimum number of balls of radius η needed to cover X, Eq. (4.3) yields

$$\mathcal{N}(H, \varepsilon) \leqslant \mathcal{N}(\Sigma(H), \varepsilon) \leqslant \mathcal{N}(\Omega + \Omega', \varepsilon).$$

From this bound with $\eta = \sqrt{2\varepsilon}$ follows the following upper bound on the volume of H_ε:

$$\begin{aligned}\mathcal{H}^d \times \mathcal{H}^d(H_\varepsilon) &\leqslant \mathcal{H}^d \times \mathcal{H}^d(H^{\sqrt{2\varepsilon}})\\ &\leqslant \mathrm{const}(d)\mathcal{N}(\Omega + \Omega', \sqrt{\varepsilon})\varepsilon^d\\ &\leqslant \mathrm{const}(d)\mathcal{H}^d(\Omega + \Omega')\varepsilon^{d/2}\end{aligned}$$

The last inequality follows from the usual comparison between covering and packing numbers. □

Note that there are simple examples for which the exponent of ε in the Lemma is sharp. Consider $\Omega = \Omega' = \mathrm{B}(0, 1)$ and $\psi(x) = \frac{1}{2}\|x\|^2$. A simple

computation shows that the ε-subdifferential of the function ψ at x is the ball $\mathrm{B}(x, \sqrt{2\varepsilon})$. Hence,

$$\mathrm{Graph}(\partial^\varepsilon \psi) = \{(x, y); x \in \Omega, y \in \mathrm{B}(x, \sqrt{2\varepsilon})\}.$$

This implies that the Hausdorff measure $\mathcal{H}^{2d}(\mathrm{Graph}(\partial^\varepsilon \psi))$ is lower bounded by a constant depending on the dimension multiplied by $\varepsilon^{d/2}$.

4.2. **Proof of Theorem 4.1.** We consider the convex function

$$\psi : x \mapsto \frac{1}{2}\|x\|^2 - \min_{y \in Y}\left(\frac{1}{2}\|x - y\|^2 + w(y)\right).$$

As a consequence of Lemma 4.2, we only need to bound the cardinal of the intersection $\mathrm{Graph}(\partial^\varepsilon \psi) \cap (X \times Y)$. Denote by χ_x^η the uniform probability measure on the ball of radius $\eta := \sqrt{\varepsilon}$ centered at x, and $\mu_\eta = \frac{1}{N}\sum_{x \in X} \chi_x^\eta$, $\nu_\eta = \frac{1}{N}\sum_{y \in Y} \chi_y^\eta$. The density μ_η is upper bounded by $\frac{n}{N\beta_d \eta^d}$ on Ω, where β_d denotes the d-volume of the unit ball. Morever, by the lower bound on the cardinal of the intersection $|B(x, \eta) \cap X|$, we know that $N\beta_d\eta^d$ is larger than the volume of Ω. Similar bounds hold for ν. Hence, denoting by G_ε the set $\mathrm{Graph}(\partial^\varepsilon \psi)$, we have

$$\begin{aligned}
\frac{1}{N^2}|G_\varepsilon \cap (X \times Y)| &\leqslant [\mu_\eta \times \nu_\eta](G_\varepsilon^{2\eta}) \\
&\leqslant \frac{nm}{\mathcal{H}^d(\Omega)\mathcal{H}^d(\Omega')}[\mathcal{H}^d \times \mathcal{H}^d](G_0^{\sqrt{\varepsilon}+2\eta}) \\
&\leqslant \mathrm{const}(d) nm \frac{\mathcal{H}^d(\Omega + \Omega')}{\mathcal{H}^d(\Omega)\mathcal{H}^d(\Omega')}\varepsilon^{d/2}
\end{aligned}$$

We conclude using $N\beta_d\eta^d \leqslant n\mathcal{H}^d(\Omega)$ and $N\beta_d\eta^d \leqslant m\mathcal{H}^d(\Omega')$.

5. Two applications in geometry

5.1. **Discretization of optimal transport.** The discrete optimal transport problem can be generalized as follows. Given two probability measures μ and ν on two compact metric spaces A and B, a *transport plan* between μ and ν is a probability measure π on the product space $X \times Y$ whose marginal on A and B is respectively μ and ν, or said otherwise,

$$\forall R \subseteq A,\ \pi(R \times B) = \mu(A) \text{ and } \forall S \subseteq B,\ \pi(A \times S) = \nu(B)$$

Again, the set of optimal transport plans is denoted by $\Gamma(\mu, \nu)$. Given a cost function $c : X \times Y \to \mathbb{R}$, the *optimal transport problem* between μ and ν consists in the following optimization problem:

$$\min\left\{\int_{X \times Y} c(x, y)\mathrm{d}\pi(x, y);\ \pi \in \Gamma(\mu, \nu)\right\}.$$

In the following, we will approximate μ and ν by two discrete measures $\overline{\mu}$ and $\overline{\nu}$ that can be written as finite sums of unit Dirac masses. This discretization is meaningful because of the stability of optimal transport plans [27, Theorem 5.20]. This theorem asserts that, as $\overline{\mu}$ and $\overline{\nu}$ weakly converge to μ and ν, the optimal transport plan $\overline{\pi}$ between $\overline{\mu}$ and $\overline{\nu}$ weakly converges to a solution of the optimal transport problem between μ and ν. It should be noted, however, that there is no known quantitative versions of this stability results even in simple cases.

5.2. **Quantization of a measure on a convex surface.** In the applications that we consider below, we need to discretize a probability measure with density on a convex surface by a sum of unit Dirac masses. In this paragraph, we describe a simple approach that constructs relevant approximation of a given probability measure with density by an average of k unit Dirac masses, which converges weakly to the target measure when the size of the sampling goes to infinity.

Optimal quantization. Given a probability measure with density $f : \partial K \to \mathbb{R}$ on the boundary of a convex set K, and a point set $P \subseteq \partial P$, the minimum quadratic Wasserstein distance (with respect to the intrinsic distance on ∂P) between f and a discrete probability measure supported on P is measured by the *quantization error* :

$$E_f(P) = \int_{\partial K} \mathrm{d}_{\partial K}(x, \mathrm{p}_P(x))^2 f(x)\mathrm{d}x \tag{5.4}$$

where $\mathrm{d}_{\partial K}$ is the geodesic distance on ∂K and $\mathrm{p}_P(x)$ denotes (one of) the nearest point to x in P. The following theorem [14, 16] shows that optimal quantization can be used to construct approximations of a measure by finite sums of unit Dirac masses:

THEOREM 5.1. *Consider a probability density $f : \partial K \to [\varepsilon, \varepsilon^{-1}]$ ($\varepsilon > 0$), on the boundary of a smooth convex set, with $\dim(\partial K) = \ell$. For any $k > 0$, let*

$$P_k \in \arg\min\{E_f(P);\ \mathrm{Card}(P) = k\}$$

and let $\mu_k = \frac{1}{k}\sum_{p\in P_k} \delta_p$. Then, μ_k weakly converges to the probability measure with density $\alpha f^{\ell/(\ell+2)}$, where α is a suitable scaling constant.

Lloyd's algorithm. The non-convexity of the quantization error makes it impossible in practice to find the best point sampling P with a fixed number of points k, even for modest values of this parameter. Lloyd's algorithm[20], which is a standard technique in geometry and image processing, is an easy to implement fixed point strategy that finds a reasonable critical point of this quantization error. This algorithm is described in Algorithm 2.

Algorithm 2 Continuous Lloyd's algorithm

Step 0: Construct P by sampling k distinct random points from μ.

Step 1: Compute the Voronoi diagram of P, move every point p of P to the centroid of the Voronoi cell with respect to f, i.e.

$$p \leftarrow \arg\min_q \int_{\mathrm{Vor}_P(p)} d(x,q)^2 f(x)\mathrm{d}x, \text{ where } \mathrm{Vor}_P(p) = \mathrm{p}_P^{-1}(\{p\}).$$

If the mass of the Voronoi cell vanishes, the point p does not move.

Step 2: Continue while all points move by less than a threshold in Step 1.

In practice, one can replace the computation of the surface Voronoi cell for the intrinsic metric $\mathrm{Vor}_P(p)$ by the intersection of the 3D Voronoi cell with the tangent plane to ∂K at p, and similarly for the construction of the weighted centroid. The centroid then needs to be projected back on the surface. Note that it would be possible to replace the linearly-converging Lloyd

algorithm by a quasi-Newton algorithm, with almost quadratic convergence speed, following [19].

5.3. **Regularity of optimal transport on oblate ellipsoids.** The question of regularity of optimal transport plans has received a lot of attention. Examples of discontinuous transport plans have been found by Caffarelli, between probability densities on the plane with non-convex support, and by Loeper on compact manifolds with negative curvature even for densities lower bounded by a positive constant (see e.g. Chapter 12 in [27]). An even more striking recent result is that even on simple convex surfaces, it is possible to find examples of pairs of non-vanishing probability densities such that the optimal transport map is discontinuous. In this section we study numerically by our optimization method the case of oblate ellipsoids, which is one of the simplest cases where discontinuity can occur [7, 13].

The family of ellipsoids (S_μ) we will consider is defined by

$$S_\mu := \left\{ (x, y, z) \in \mathbb{R}^3;\ x^2 + y^2 + \left(\frac{z}{\mu}\right)^2 = 1 \right\} \tag{5.5}$$

Bonnard, Caillau and Rifford [7] have shown that for μ smaller than $1/\sqrt{3}$, there exists a point on the equator $\{z = 0\}$ whose injectivity domain is non-convex. By [13], this fact implies the existence of two absolutely continuous probability measure μ, ν with densities in $[C^{-1}, C]$ such that the optimal transport plan between μ and ν is discontinuous.

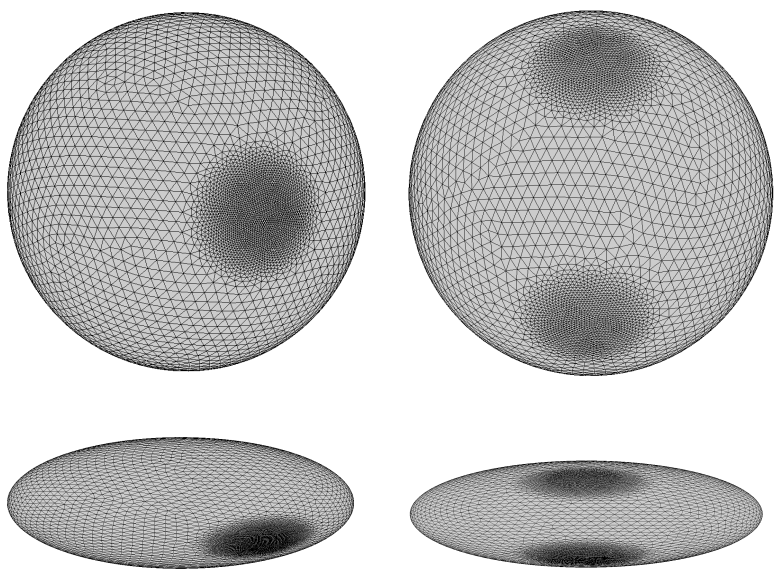

FIGURE 1. Quantization by 10^4 Dirac masses of the initial and the target measures on the sphere (first row) and on the oblate ellipsoid $S_{0.3}$ (second row)

To obtain a qualitative description of this phenomena, we introduce two non-vanishing densities with equal total mass on S_1, whose formula in Cartesian coordinates is given by:

$$f(x,y,z) = 0.05 + 4\exp\left(-c(1-x)^2\right)$$

and

$$g(x,y,z) = 0.05 + 2\exp\left(-c(1-\frac{\sqrt{2}}{2}(z+y))^2\right) + 2\exp\left(-c(1-\frac{\sqrt{2}}{2}(z-y))^2\right)$$

with $c = 500$. Previous parameters have been selected to define non-vanishing densities which are not too concentrated. Even in this regular situation, we will see that optimal transport maps may have very different qualitative behaviors depending on the eccentricity of the ellipsoid.

We construct a quantization by a sum of 10^4 unit Dirac masses of the two measures of densities f and g with respect to the two dimensional Hausdorff measure on S_1 using Lloyd's algorithm, as described in §5.2. Our samplings are reported in figure 1. Our goal is to compare the optimal transport between the two discrete measures on the sphere and between two analogous measures defined on the oblate ellipsoid S_μ, with $\mu = 0.3$. By analogous we mean that we deduce our two measures on S_μ by the naive bi-Lipschitz transformation

$$(x,y,z) \to (x,y,\mu z).$$

Implementation. There is no known analytical formula to compute the geodesic distance between two points on an ellipsoid which is not a sphere. In order to compute a precise approximation of the cost matrix we used a software library developed by Krisanov, which is based on a continuous version of Djikstra's algorithm described in [26]. Let us point out that this algorithm is exact on a triangular surface and has a complexity of $N^2 \log N$ where N stands for the number of points on the surface. The optimal transport plans between both pair of measures are then obtained simply by applying the algorithm described in §3 to the matrix of squared geodesic distances. This takes a few minutes on a standard computer.

To illustrate the difference between the two optimal transports, we computed for every sampling point of the source densities the geodesic distance between its image and the center of the Gaussian part of the source density, i.e. the point $c := (1,0,0)$. More formally, this is equivalent to compute the pullback of the geodesic distance function to $(1,0,0)$ by the optimal transport. These functions are rescaled to have the same infinity norm, and their values are reported in Figure 2, with the same color map. One can observe that the function is smooth along the equator in the spherical case ($\mu = 1$) whereas in the oblate case ($\mu = 0.3$) some points close to the equator can be transported to points whose geodesic distance to c is very different. This qualitative behavior, observed here at a discrete level, is the first numerical illustration of the discontinuity of optimal transport plans between two non-vanishing densities supported on a convex manifold.

5.4. **Alexandrov's reconstruction problem.** In this section, we propose an method to reconstruct a 3D convex body from its Gaussian curvature

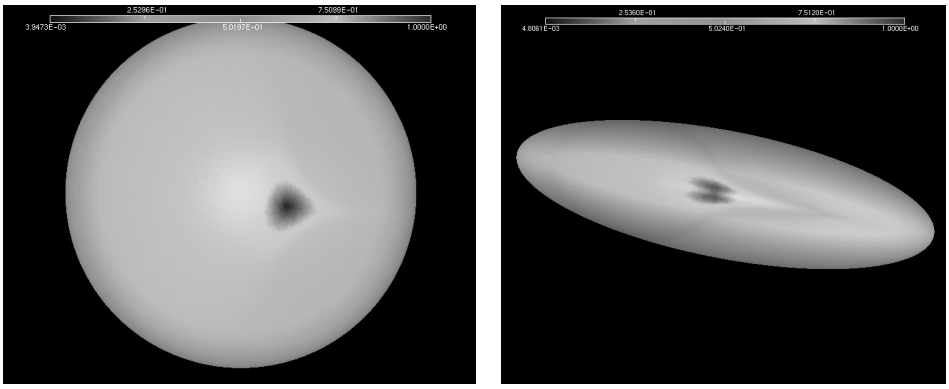

FIGURE 2. Pullback of the geodesic distance function to $(1, 0, 0)$ by the optimal transports on the sphere and on the oblate ellipsoid respectively.

measure based on a variational characterisation of [23, 5] and on our discrete optimal transport algorithm.

Optimal transport formulation of Alexandrov's theorem. Let $K \subseteq \mathbb{R}^d$ be a convex body in $\mathbb{R}^d$, containing the origin in its interior. Any convex set admits an exterior unit normal vector field $\mathbf{n}_K : \partial K \to \mathcal{S}^{d-1}$, which is uniquely defined almost everywhere. Let σ be the probability measure on the unit sphere, obtained by rescaling the $(d-1)$-dimensional Hausdorff measure. The *Gaussian measure* $\mathcal{G}_K$ of K is by definition the pullback of σ, by the Gauss map $\mathbf{n}_K$. More explicitely,

$$\forall B \subseteq \partial K,\ \mathcal{G}_K(B) := \sigma(\mathbf{n}_K(B))$$

Since K contains the origin in its interior, its boundary can be parameterized by a radial map $\rho_K : \mathcal{S}^{d-1} \to \partial K$. For every direction u in $\mathcal{S}^{d-1}$, $\rho_K(u)$ lies in the intersection of ∂K with the ray $\{tu; t \geqslant 0\}$. We can again pull-back the measure $\mathcal{G}_K$ by the map ρ_K, thus defining a measure on the unit sphere $\mathcal{G}_K^0$, which we will call *Alexandrov measure.*

$$\forall B \subseteq \mathcal{S}^{d-1},\ \mathcal{G}_K^0(B) := \sigma(\mathbf{n}_K \circ \rho_K(B)).$$

Alexandrov addressed in [1] the question of the existence and uniqueness (up to homotethy) of a convex body with prescribed Alexandrov measure μ, under some conditions on μ. The relationship between this reconstruction problem and a problem of optimal transport on the unit sphere for the cost $c(u, v) = -\log(\max(\langle u|v\rangle, 0))$ has been first remarked by Oliker [23], and then used by Bertrand to give a direct variational proof of Alexandrov theorem [5]. Bertrand's version of Alexandrov's theorem says the following:

THEOREM 5.2. *Given a probability measure μ on the unit sphere, there exists a convex body K such that $\mathcal{G}_K^0 = \mu$ if and only the following optimal transport problem between σ and μ for the cost function $c(u, v) := -\log(\max(\langle u|v\rangle, 0))$ admits a solution with finite cost:*

$$\sup_{\varphi,\psi} \int_{\mathcal{S}^{d-1}} \varphi(u)\mathrm{d}\sigma - \int_{\mathcal{S}^{d-1}} \psi(v)\mathrm{d}\mu = \inf_{\pi} \int_{\mathcal{S}^{d-1}\times\mathcal{S}^{d-1}} c(u, v)\mathrm{d}\pi(u, v) \qquad (5.6)$$

where the maximum is taken over functions φ, ψ satisfying the relation $\varphi(u) - \psi(v) \leqslant c(u,v)$, and the infimum is taken over transport plans between u and v.

Moreover, from Section 2.2 of [5], one can see that if from a pair φ, ψ maximizing the left-hand side, one can reconstruct the radial function of a convex set solving Alexandrov's problem by $\rho_K(u) = \exp(-\psi(u))u$. Note that one could also reconstruct the support function of K from φ.

Example: Perturbed icosahedron. Theorem 4.1 suggests that concentrated measures can be a source of numerical difficulties. In order to evaluate the reconstruction procedure and the efficiency of our algorithm in highly concentrated situations we consider the following test case. Let (V_i) be the vertices of one of the five Platonic solid inscribed in the unit sphere. Since these polytopes are regular, the associated Gaussian curvature is a sum of Dirac masses at each V_i with mass equal to 4π over the number of vertices. To create an concentrated measure we sample several points from each V_i and perturb them randomly. More explicitely, we consider the measure

$$\mu = c \sum_i^n \sum_j^m \delta_{R^j(V_i,s)},$$

where $R^1(V_i,s), \ldots, R^m(V_i,s)$ are obtained by taking m independent samples of the isotropic Gaussian vector centered at V_i with variance σ, and projecting them on the unit sphere. The constant c is adjusted in order to give a total mass of 4π. We carried out this experiment in the case of the icosahedron so that $n = 12$ and we fixed $m = 50$ and $s = 10^{-4}$. We used a quantization of the uniform measure σ on the unit sphere by a sum of nm Dirac masses using Lloyd's algorithm, as described in §5.2.

The algorithm stops after 33 steps of the local linearization. The reconstructed polytope and a zoom at a vertex are displayed in Figure 4.

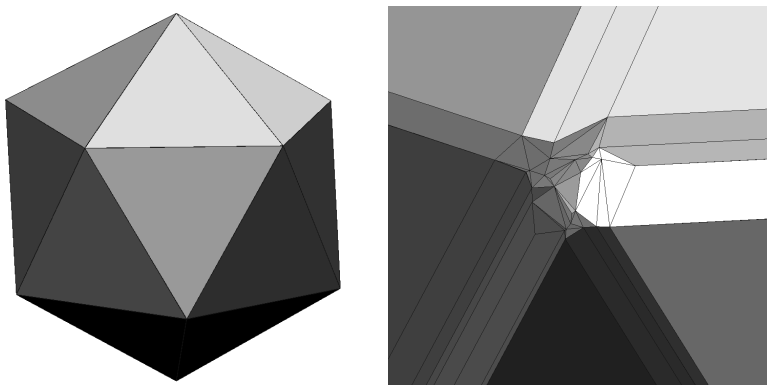

FIGURE 3. Reconstruction of a random perturbed icosahedron by its Gauss curvature measure

Example: Convex set with 1*-dimensional Gaussian measure.* Finally to illustrate the reconstruction of a convex body which is not a polytope, we consider the following Gaussian measure

$$\mu = c \sum_{i=1}^{3} \mathcal{H}^1_{|C_i},$$

where $\mathcal{H}_{|C_i}$ stands for the restriction of the one-dimensional uniform Hausdorff measure to a circular arc C_i. We fixed C_1 and C_2 to be to circles of same radius contained in horizontal plans at altitude $z = 0.5$ and $z = -0.5$ respectively, and we defined C_3 to be the great circle contained in the plane $x = 0$. As in the previous example, we adjust the constant c so as to give a total mass of 4π. To obtain a discrete optimal transport problem problem, we discretize the measure μ by sampling 300 uniform points on each circle. Only two local linearization are needed to reach convergence.

The resulting points on the boundary of the convex body and the convex hull of those vertices are shown in Figure 4.

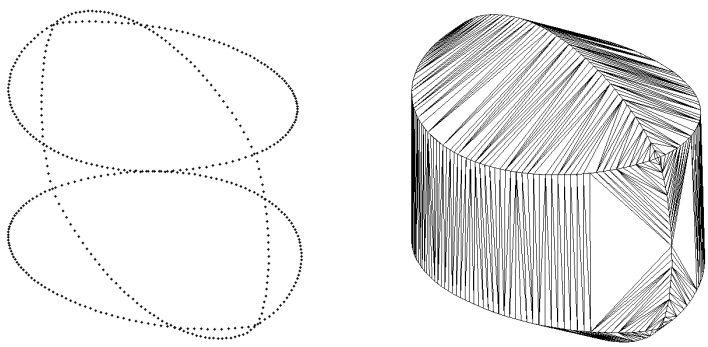

FIGURE 4. Reconstruction of a convex body described by its Gaussian curvature

REFERENCES

1. A.D. Aleksandrov, *Convex polyhedra*, Springer Verlag, 2005.
2. F. Aurenhammer, F. Hoffmann, and B. Aronov, *Minkowski-type theorems and least-squares clustering*, Algorithmica **20** (1998), no. 1, 61–76.
3. J.D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the monge-kantorovich mass transfer problem*, Numerische Mathematik **84** (2000), no. 3, 375–393.
4. Jean-David Benamou, Brittany Froese, and Adam Oberman, *Numerical solution of the second boundary value problem for the Elliptic Monge-Ampère equation*, Rapport de recherche, 2012.
5. J. Bertrand, *Prescription of gauss curvature using optimal mass transport*, (2010).
6. D.P. Bertsekas and J. Eckstein, *Dual coordinate step methods for linear network flow problems*, Mathematical Programming **42** (1988), no. 1, 203–243.
7. B. Bonnard, J.B. Caillau, and L. Rifford, *Convexity of injectivity domains on the ellipsoid of revolution: The oblate case*, Comptes Rendus Mathematique **348** (2010), no. 23, 1315–1318.
8. A. Brondsted and RT Rockafellar, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc, vol. 16, 1965, pp. 605–611.

9. R.E. Burkard, M. Dell'Amico, and S. Martello, *Assignment problems*, Society for Industrial Mathematics, 2009.
10. L. Buš and P. Tvrdík, *Towards auction algorithms for large dense assignment problems*, Computational Optimization and Applications **43** (2009), no. 3, 411–436.
11. LA Caffarelli and VI Oliker, *Weak solutions of one inverse problem in geometric optics*, Journal of Mathematical Sciences **154** (2008), no. 1, 39–49.
12. MJP Cullen and RJ Purser, *An extended lagrangian theory of semi-geostrophic frontogenesis.*, Journal of Atmospheric Sciences **41** (1984), 1477–1497.
13. A. Figalli, L. Rifford, and C. Villani, *Necessary and sufficient conditions for continuity of optimal transport maps on riemannian manifolds*, Tohoku Mathematical Journal **63** (2011), no. 4, 855–876.
14. P.M. Gruber, *Optimum quantization and its applications*, Advances in Mathematics **186** (2004), no. 2, 456–497.
15. J.B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, vol. 1, Springer, 1996.
16. B. Kloeckner, *Approximation by finitely supported measures*, ESAIM: Control, Optimisation and Calculus of Variations (2011), To appear.
17. A.S. Lewis and M.L. Overton, *Behavior of BFGS with an exact line search on nonsmooth examples*, Tech. report, Technical report, Optimization Online, 2008b. http://www. optimization-online. org/DB_FILE/2008/12/2173. pdf, submitted to SIAM J. Optimization, 2008.
18. ———, *Nonsmooth optimization via BFGS*, SIAM Journal of Optimization, submitted for publication (2009).
19. Y. Liu, W. Wang, B. Lévy, F. Sun, D.M. Yan, L. Lu, and C. Yang, *On centroidal voronoi tessellation—energy smoothness and fast computation*, ACM Transactions on Graphics **28** (2009), no. 4, 101.
20. S. Lloyd, *Least squares quantization in pcm*, Information Theory, IEEE Transactions on **28** (1982), no. 2, 129–137.
21. G. Loeper and F. Rapetti, *Numerical solution of the monge–ampère equation by a newton's algorithm*, Comptes Rendus Mathematique **340** (2005), no. 4, 319–324.
22. Q. Mérigot, *A multiscale approach to optimal transport*, **30** (2011), no. 5, 1583–1592.
23. V. Oliker, *Embedding $\mathcal{S}^n$ into $\mathbb{R}^{n+1}$ with given integral Gauss curvature and optimal mass transport on $\mathcal{S}^n$*, Advances in Mathematics **213** (2007), no. 2, 600–620.
24. VI Oliker and LD Prussner, *On the numerical solution of the equation* $\frac{\partial^2 z}{\partial x}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y}\right)^2 = f$, Numerische Mathematik **54** (1988), no. 3, 271–293.
25. A.V. Pogorelov, *Monge–Ampère equations of elliptic type*, (1964).
26. V. Surazhsky, T. Surazhsky, D. Kirsanov, S.J. Gortler, and H. Hoppe, *Fast exact and approximate geodesics on meshes*, ACM Transactions on Graphics (TOG), vol. 24, ACM, 2005, pp. 553–560.
27. C. Villani, *Optimal transport: old and new*, vol. 338, Springer Verlag, 2009.