J. Eur. Math. Soc. 21, 2603–2651
DOI 10.4171/JEMS/889

Jun Kitagawa · Quentin Mérigot · Boris Thibert

Convergence of a Newton algorithm for semi-discrete optimal transport

Received April 1, 2016

Abstract. A popular way to solve optimal transport problems numerically is to assume that the source probability measure is absolutely continuous while the target measure is finitely supported. We introduce a damped Newton algorithm in this setting, which is experimentally efficient, and we establish its global linear convergence for cost functions satisfying an assumption that appears in the regularity theory for optimal transport.

Keywords. Optimal transport, Ma–Trudinger–Wang condition, Laguerre tessellation

Contents

1. Introduction

Some problems in geometric optics or convex geometry can be recast as optimal transport problems between probability measures: this includes the far-field reflector antenna problem, Aleksandrov's Gaussian curvature prescription problem, etc. A popular way to solve these problems numerically is to assume that the source probability measure is

J. Kitagawa: Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA; e-mail: kitagawa@math.msu.edu

Q. Mérigot: Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, F-91405 Orsay, France; e-mail: quentin.merigot@math.u-psud.fr

B. Thibert: Laboratoire Jean Kuntzmann, Université Grenoble Alpes, CNRS, F-38000 Grenoble, France; e-mail: boris.thibert@univ-grenoble-alpes.fr

Mathematics Subject Classification (2010): Primary 49M25; Secondary 65K10

absolutely continuous while the target measure is finitely supported. We refer to this setting as semi-discrete optimal transport. Among the several algorithms proposed to solve semi-discrete optimal transport problems, one currently needs to choose between algorithms that are slow but come with a convergence speed analysis [29, 8, 21] or algorithms that are much faster in practice but which come with no convergence guarantees [5, 27, 11, 22, 10]. Algorithms of the first kind rely on coordinatewise increments, and the number of iterations required to reach the solution up to an error of ε is of order N^3/ε, where N is the number of Dirac masses in the target measure. On the other hand, algorithms of the second kind typically rely on the formulation of the semi-discrete optimal transport problem as an unconstrained convex optimization problem which is solved using a Newton or quasi-Newton method.

The purpose of this article is to bridge this gap between theory and practice by introducing a damped Newton algorithm which is experimentally efficient and by proving the global convergence of this algorithm with optimal rates. The main assumption is that the cost function satisfies a condition that appears in the regularity theory for optimal transport (the Ma–Trudinger–Wang condition) and that the support of the source density is connected in a quantitative way (it must satisfy a weighted Poincaré–Wirtinger inequality). In §1.7, we compare this algorithm and the convergence theorem to previous computational approaches to optimal transport.

1.1. Semi-discrete optimal transport

The source space is an open domain Ω of a d-dimensional Riemannian manifold, which we endow with the measure $\mathcal{H}^d_g$ induced by the Riemannian metric g on the manifold. The target space is an (abstract) finite set Y. We are given a cost function c on the product space $\Omega \times Y$, or equivalently a collection $(c(\cdot, y))_{y\in Y}$ of functions on Ω. We assume that the functions $c(\cdot, y)$ are of class $\mathcal{C}^{1,1}$ on Ω:

$$\forall y \in Y, \qquad c(\cdot, y) \in \mathcal{C}^{1,1}(\Omega). \tag{Reg}$$

Here $\mathcal{C}^{n,\alpha}(\Omega)$ denotes the class of functions which are n-times differentiable and whose n-th derivatives are α-Hölder continuous. In particular, $\mathcal{C}^{0,\alpha}$ is the space of α-Hölder continuous functions. We consider a compact subset X of Ω and a probability density ρ on X, i.e. $\rho\, \mathrm{d}\mathcal{H}^d$ is a probability measure. By an abuse of notation, we will often conflate the density ρ with the measure $\rho\, \mathrm{d}\mathcal{H}^d$ itself. Note that the support of ρ is contained in X, but we do not assume that it is actually equal to X. The push-forward of ρ by a measurable map $T : X \to Y$ is the finitely supported measure $T_\# \rho = \sum_{y\in Y} \rho(T^{-1}(y))\delta_y$. The map T is called a *transport map* between ρ and a probability measure μ on Y if $T_\#\rho = \mu$. The *semi-discrete optimal transport problem* consists in minimizing the transport cost over all transport maps between ρ and μ, that is,

$$\min\left\{ \int_X c(x, T(x))\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x) \;\middle|\; T : X \to Y \text{ such that } T_\#\rho = \mu \right\}. \tag{M}$$

This problem is an instance of Monge's optimal transport problem, where the target measure is finitely supported. Kantorovich proposed a relaxed version of the problem (M) as an infinite-dimensional linear programming problem over the space of probability measures with marginals ρ and μ.

1.2. Laguerre tessellation and economic interpretation

In the semi-discrete setting, the dual of Kantorovich's relaxation can be conveniently phrased using the notion of Laguerre tessellation. We start with an economic metaphor. Assume that the probability density ρ describes the population distribution over a large city X, and that the finite set Y describes the location of bakeries in the city. Customers living at a location x in X try to minimize the walking cost $c(x, y)$, resulting in a decomposition of the space called a Voronoi tessellation. The number of customers received by a bakery $y \in Y$ is equal to the integral of ρ over its Voronoi cell, namely

$$\mathrm{Vor}_y := \{x \in \Omega \mid \forall z \in Y,\ c(x, y) \leq c(x, z)\}.$$

If the price of bread is given by a function $\psi : Y \to \mathbb{R}$, customers living at location x in X make a compromise between walking cost and price by minimizing the sum $c(x, y)+\psi(y)$. This leads to the notion of Laguerre tessellation, whose cells are given by

$$\mathrm{Lag}_y(\psi) := \{x \in \Omega \mid \forall z \in Y,\ c(x, y) + \psi(y) \leq c(x, z) + \psi(z)\}. \tag{1.1}$$

When the sets X and Y are contained in $\mathbb{R}^d$ and the cost is the squared Euclidean distance, the computation of the Laguerre tessellation is a classical problem of computational geometry, for which there exists very efficient software, such as CGAL [1] or Geogram [2]. For other cost functions, one has to adapt the algorithms, as was done for the reflector cost on the sphere in [10]. The shape of the Voronoi and Laguerre tessellations is depicted in Figure 1.

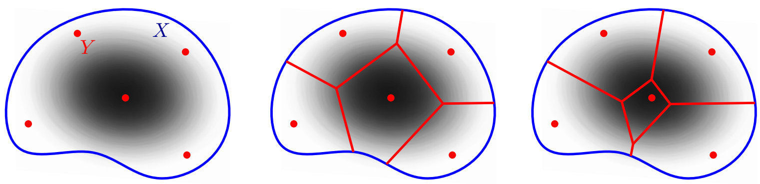

Fig. 1. Left: the domain X (with boundary in blue) is endowed with a probability density pictured in grayscale representing the density of population in a city. The set Y (in red) represents the location of bakeries. Here, $X, Y \subseteq \mathbb{R}^2$ and $c(x, y) = |x - y|^2$. Middle: The Voronoi tessellation induced by the bakeries. Right: The Laguerre tessellation: the price of bread at the bakery near the center of X is higher than at the other bakeries, effectively shrinking its Laguerre cell.

We want the Laguerre cells to form a partition of Ω up to a negligible set. By the implicit function theorem, this will be the case if the following *twist condition* holds:

$$\forall x \in X, \qquad Y \ni y \mapsto \mathrm{D}_x c(x, y) \in T_x^*\Omega \text{ is injective,} \tag{Twist}$$

where D_x denotes differentiation with respect to the first variable. The twist condition implies that for any prices ψ on Y, the transport map induced by the Laguerre tessellation,

$$T_\psi(x) := \arg\min_{y \in Y}(c(x, y) + \psi(y)), \tag{1.2}$$

is uniquely defined almost everywhere. It is easy to see (Proposition 2.2) that for any function ψ on Y, the map T_ψ is an optimal transport map between ρ and the push-forward measure $T_{\psi\#}\rho = \sum_{y\in Y} \rho(\mathrm{Lag}_y(\psi))\delta_y$.

1.3. *Kantorovich's functional*

The map T_ψ is an optimal transport map between ρ and $T_{\psi\#}\rho$. Conversely, Theorem 1.1 below ensures that any semi-discrete optimal transport problem admits such a solution. In other words, for any probability density ρ on X and any probability measures μ on Y there exists a function (price) ψ on Y such that $T_{\psi\#}\rho = \nu$. The proof of this theorem is an easy generalization of the proof given in [5] for the quadratic cost, but it is nonetheless included in Section 2 for the sake of completeness.

Here and below, we denote by $(\mathbb{1}_y)_{y\in Y}$ the canonical basis of $\mathbb{R}^Y$, and by $\|\cdot\|$ the Euclidean norm induced by this basis, while $\|\cdot\|_g$ will denote the norm induced by the Riemannian metric g on either $T_x\Omega$ or $T_x^*\Omega$ (which will be clear from context). We will, slightly abusively, consider the space $\mathcal{P}(Y)$ of probability measures as a subset of $\mathbb{R}^Y$.

Theorem 1.1. *Assume* (Reg) *and* (Twist)*, and let ρ be a bounded probability density on X and $\nu = \sum_{y\in Y} \nu_y \mathbb{1}_y$ in $\mathcal{P}(Y)$. Then the functional Φ given by*

$$\begin{aligned}\Phi(\psi) &:= \int_X \Big(\min_{y\in Y} c(x, y) + \psi(y)\Big)\rho(x)\,\mathrm{d}\mathcal{H}_g^d(x) - \sum_{y\in Y} \psi(y)\nu_y \\ &= \sum_{y\in Y} \int_{\mathrm{Lag}_y(\psi)} (c(x, y) + \psi(y))\rho(x)\,\mathrm{d}\mathcal{H}_g^d(x) - \sum_{y\in Y} \psi(y)\nu_y \qquad (1.3)\end{aligned}$$

is concave, $\mathcal{C}^1$-smooth, and its gradient is

$$\nabla\Phi(\psi) = \sum_{y\in Y} (\rho(\mathrm{Lag}_y(\psi)) - \nu_y)\mathbb{1}_y. \qquad (1.4)$$

Corollary 1.2. *The following statements are equivalent:*

(i) *$\psi : Y \to \mathbb{R}$ is a global maximizer of Φ;*
(ii) *T_ψ is an optimal transport map between ρ and ν;*
(iii) *$T_{\psi\#}\rho = \nu$, or equivalently*

$$\forall y \in Y, \quad \rho(\mathrm{Lag}_y(\psi)) = \nu_y \qquad \text{(MA)}$$

We call the function Φ introduced in (1.3) *Kantorovich's functional*. Note that both this functional and its gradient are invariant by addition of a constant. The non-linear equation (MA) can be considered as a discrete version of the generalized Monge–Ampère equation that characterizes the solutions to optimal transport problems (see for instance [34, Chapter 12]).

1.4. *Damped Newton algorithm*

We consider a simple damped Newton algorithm to solve the semi-discrete optimal transport problem. This algorithm is very close to the one used by Mirebeau [28]. To phrase this algorithm in a more general way, we introduce a notation for the measure of Laguerre cells: for $\psi \in \mathbb{R}^Y$ we set

$$G(\psi) := \sum_{y \in Y} G_y(\psi) \mathbb{1}_y \quad \text{where} \quad G_y(\psi) = \rho(\text{Lag}_y(\psi)), \tag{1.5}$$

so that $\nabla\Phi(\psi) = G(\psi) - \nu$. In Algorithm 1 below, we denote by A^+ the *pseudo-inverse* of the matrix A.

Algorithm 1. Simple damped Newton's algorithm

Input: A tolerance $\eta > 0$ and an initial $\psi_0 \in \mathbb{R}^Y$ such that

$$\varepsilon_0 := \frac{1}{2} \min\Big[\min_{y \in Y} G_y(\psi_0), \ \min_{y \in Y} \mu_y\Big] > 0. \tag{1.6}$$

While: $\|G_y(\psi_k) - \mu_y\| \geq \eta$

Step 1: Compute $d_k = -\mathrm{D}G(\psi_k)^+(G(\psi_k) - \mu)$

Step 2: Determine the minimum $\ell \in \mathbb{N}$ such that $\psi_k^\ell := \psi_k + 2^{-\ell} d_k$ satisfies

$$\begin{cases} \min_{y \in Y} G_y(\psi_k^\ell) \geq \varepsilon_0 \\ \|G(\psi_k^\ell) - \mu\| \leq (1 - 2^{-(\ell+1)}) \|G(\psi_k) - \mu\| \end{cases}$$

Step 3: Set $\psi_{k+1} = \psi_k + 2^{-\ell} d_k$ and $k \leftarrow k + 1$.

The goal of this article is to prove the global convergence of this damped Newton algorithm and to establish estimates on the speed of convergence. As shown in Proposition 6.1, the convergence of Algorithm 1 depends on the regularity and strong monotonicity of the map $G = \nabla\Phi$. As we will see, the regularity of G will depend mostly on the geometry of the cost function and the regularity of the density. On the other hand, the strong monotonicity of G will require a strong connectedness assumption on the support of ρ, in the form of a weighted Poincaré–Wirtinger inequality. Before stating our main theorem we give some indication about these intermediate regularity and monotonicity results and their assumptions.

1.5. Regularity of Kantorovich's functional and MTW condition

In order to establish the convergence of a damped Newton algorithm for (MA), we need to study the $\mathcal{C}^{2,\alpha}$ regularity of Kantorovich's functional Φ. However, while $\mathcal{C}^1$ regularity of Φ follows rather easily from the (Twist) hypothesis (or even from a weaker hypothesis, see Theorem 2.1), higher order regularity seems to depend on the geometry of the cost function in a more subtle manner. We found that a sufficient condition for the regularity of Φ is the Ma–Trudinger–Wang condition [26], which appeared naturally in the study of the regularity of optimal transport maps. We use a discretization of Loeper's geometric reformulation of the Ma–Trudinger–Wang condition [23].

Definition 1.1 (Loeper's condition). The cost c satisfies *Loeper's condition* if for every y in Y there exists a convex open subset Ω_y of $\mathbb{R}^d$ and a $\mathcal{C}^{1,1}$ diffeomorphism $\exp_y^c : \Omega_y \to \Omega$ such that the functions

$$\Omega_y \ni p \mapsto c(\exp_y^c p, y) - c(\exp_y^c p, z) \text{ are quasi-convex for all } z \text{ in } Y. \tag{QC}$$

The map $\exp_y^c$ is called the *c-exponential* with respect to y, and the domain Ω_y is an *exponential chart*.

We comment here that when Y is a finite subset of a continuous space and c satisfies conditions (Reg), (Twist), and (A3w) (see pages 2604, 2605, and 2631 respectively), the c-exponential map defined in the usual sense in optimal transport theory (see Remarks 1.1 and 4.4) will satisfy what we call Loeper's condition above. However, it will become apparent that for our purposes what is essential is the above quasi-convexity property and not the actual definition of $\exp_y^c$. Thus we will use the notation $\exp_y^c$ even in cases when Y is not a finite subset of a continuous space.

Definition 1.2 (c-Convexity). Assuming Loeper's condition, a subset X of Ω is *c-convex with respect to a point y of Y* if its inverse image $(\exp_y^c)^{-1}(X)$ is convex. The subset X is said to be *c-convex* if it is *c-convex* with respect to every point y in Y.

Note that by assumption, the domain Ω itself is c-convex. The connection between this discrete version of Loeper's condition and the conditions used in the regularity theory for optimal transport is detailed in Remark 1.1. The (QC) condition implies the convexity of each Laguerre cell in its own exponential charts, namely $(\exp_y^c)^{-1}(\mathrm{Lag}_y(\psi))$ is convex for every y in Y. This plays a crucial role in the regularity of Kantorovich's functional.

Theorem 1.3. *Assume (Reg), (Twist), and (QC). Let X be a compact, c-convex subset of Ω and let ρ be in $\mathcal{P}^{\mathrm{ac}}(X) \cap \mathcal{C}^{0,\alpha}(X)$ for α in $(0, 1]$. Then Kantorovich's functional is of class $\mathcal{C}^{2,\alpha}_{\mathrm{loc}}$ on the set*

$$\mathcal{K}^+ := \{\psi : Y \to \mathbb{R} \mid \forall y \in Y,\ \rho(\mathrm{Lag}_y(\psi)) > 0\}, \tag{1.7}$$

and its Hessian is given by

$$\begin{aligned} \frac{\partial^2 \Phi}{\partial \mathbb{1}_y \partial \mathbb{1}_z}(\psi) &= \int_{\mathrm{Lag}_y(\psi)\cap \mathrm{Lag}_z(\psi)} \frac{\rho(x)}{\|\mathrm{D}_x c(x, y) - \mathrm{D}_x c(x, z)\|_g} \,\mathrm{d}\mathcal{H}_g^{d-1}(x) \quad (z \neq y), \\ \frac{\partial^2 \Phi}{\partial \mathbb{1}_y^2}(\psi) &= - \sum_{z \in Y\setminus\{y\}} \frac{\partial^2 \Phi}{\partial \mathbb{1}_y \partial \mathbb{1}_z}. \end{aligned} \tag{1.8}$$

The proof of this theorem and a more precise statement are given in Section 4 (Theorem 4.1), showing that the $\mathcal{C}^{2,\alpha}$ estimate can be made uniform when the mass of the Laguerre cells is bounded from below by a positive constant.

Remark 1.1. We remark that under certain assumptions on the cost c, our (QC) condition is implied by classical conditions introduced in a smooth setting by X.-N. Ma, N. Trudinger, and X.-J. Wang [26], which include the well known (MTW) or (A3) condition. See Remark 4.4 for more specifics.

There are a wide variety of known examples satisfying these conditions. Aside from the canonical example of the inner product on $\mathbb{R}^n \times \mathbb{R}^n$, and other costs on Euclidean spaces mentioned in [26, 33], there are the non-flat examples of the Riemannian distance squared and $-\log \|x - y\|_{\mathbb{R}^{n+1}}$ on (a subset of) $\mathbb{S}^n \times \mathbb{S}^n$ (see [24]). The last cost is associated to the *far-field reflector antenna problem*. We refer the reader to [19, p. 1331] for a (more) comprehensive list of such costs.

1.6. Strong concavity of Kantorovich's functional

As noted earlier, Kantorovich's functional Φ cannot be strictly concave, since it is invariant under addition of a constant. This implies that the Hessian $\mathrm{D}^2\Phi$ has a zero eigenvalue corresponding to the constants. A more serious obstruction to the strict concavity of Φ at a point ψ arises when the discrete graph induced by the Hessian (where two points are connected iff $\partial^2\Phi/\partial\mathbb{1}_y\partial\mathbb{1}_z(\psi) \neq 0$) is not connected. This can happen either because one of the Laguerre cells is empty (hence not connected to any neighbor) or if the support of the probability density ρ is itself disconnected. In order to avoid the latter phenomena, we will require that (X, ρ) satisfies a weighted L^1 Poincaré–Wirtinger inequality.

Definition 1.3 (weighted Poincaré–Wirtinger). A continuous probability density ρ on a compact set $X \subseteq \Omega$ satisfies a *weighted Poincaré–Wirtinger inequality* with constant $C_{\mathrm{pw}} > 0$ if for every $\mathcal{C}^1$ function f on X,

$$\|f - \mathbb{E}_\rho(f)\|_{L^1(\rho)} \leq C_{\mathrm{pw}}\|\nabla f\|_{L^1(\rho)}, \tag{PW}$$

where $\|h\|_{L^1(\rho)} := \int_X |h(x)|\rho(x)\,\mathrm{d}\mathcal{H}^d_g(x)$ and $\mathbb{E}_\rho(f) := \int_X f(x)\rho(x)\,\mathrm{d}\mathcal{H}^d_g(x)$.

We denote by E_Y the orthogonal complement (in $\mathbb{R}^Y$) of the space of constant functions on Y, that is, $E_Y := \{\psi \in \mathbb{R}^Y \mid \sum_y \psi(y) = 0\}$. As before, $\mathcal{K}^+$ is the set of functions ψ whose Laguerre cells all have positive mass.

Theorem 1.4. *Assume* (Reg), (Twist), *and* (QC). *Let X be a compact, c-convex subset of Ω, and ρ be a continuous probability density on X satisfying* (PW). *Then Kantorovich's functional Φ is strictly concave on $E_Y \cap \mathcal{K}^+$.*

As before, a more quantitative statement is proven in Section 5 (Theorem 5.1), establishing strong concavity of Φ under the assumption that the mass of the Laguerre cells is bounded from below by a positive constant.

1.7. Convergence result

Putting Proposition 6.1, Theorem 1.3 and Theorem 1.4 together, we can prove the global convergence of the damped Newton algorithm for semi-discrete optimal transport (Algorithm 1) together with optimal convergence rates.

Theorem 1.5. *Assume* (Reg), (Twist)*, and* (QC)*, and also that*

(i) *the support of the probability density ρ is included in a compact, c-convex subset X of Ω, and $\rho \in \mathcal{C}^{0,\alpha}(X)$ for α in* $(0, 1]$*,*
(ii) *ρ has positive Poincaré–Wirtinger constant.*

Then the damped Newton algorithm for semi-discrete optimal transport (*Algorithm* 1) *converges globally with linear rate and locally with rate $1+\alpha$.*

Remark 1.2. This theorem makes no assumption about the convexity (or c-convexity) of the support of the source density ρ. Such cases are not handled by other numerical methods for Monge–Ampère equations [6, 24]. For completeness, in Appendix A we provide an explicit example of a radial measure on $\mathbb{R}^d$ whose support is an annulus but whose Poincaré–Wirtinger constant is nonetheless positive.

Remark 1.3. The positive lower bound on the damping parameter ($\tau_k = 2^{-\ell}$ in Algorithm 1) established in this theorem degrades as N grows to infinity. It is plausible (but far from direct) that one could control this quantity when N is large by a comparison to the continuous Monge–Ampère equation. The strong concavity estimate (Theorem 1.4) would then need to be replaced by uniform ellipticity estimates for the linearized Monge–Ampère equation, while the regularity estimate (Theorem 1.3) would be replaced by regularity estimates for solutions to the Monge–Ampère equation. We refer to Loeper and Rapetti [25] for an implementation of this ideas in a continuous setting. The space-discretization of their approach is open.

Comparison to previous work. There exist a few other numerical methods relying on Newton's algorithm for the resolution of the standard Monge–Ampère equation or for the quadratic optimal transport problem. Here, we highlight some of the differences between Algorithm 1 and Theorem 1.5 and these existing results. First, we note that many authors have reported the good behavior in practice of Newton's or quasi-Newton's methods for solving discretized Monge–Ampère equations or optimal transport problems [27, 11, 6]. Note however that none of these works contain convergence proofs for the Newton algorithm.

Loeper and Rapetti [25] (their result was refined by Saumier, Agueh, and Khouider [31]) establish the global convergence of a damped Newton method for solving quadratic optimal transport on the torus, relying heavily on Caffarelli's regularity theory. In particular, the convergence of the algorithm requires a positive lower bound on the probability densities, while this condition is not necessary for Theorem 1.5 (see Section 5 and Appendix A where we explicitly construct probability densities with non-convex support that still satisfy the hypothesis of Theorem 1.5). A second drawback of relying on the regularity theory for optimal transport is that the damping parameter, which is an input parameter of the algorithm used in [25], cannot be determined explicitly from the data. Third, the convergence proof is for continuous densities, and it seems difficult to adapt it to the space-discretized problem. On the positive side, it seems likely that the convergence proof of [25], [31] can be adapted to cost functions satisfying the Ma–Trudinger–Wang condition (which is equivalent to Loeper's condition (QC) that we also require).

Oliker and Prussner [29] prove the *local* convergence of Newton's method for finding Aleksandrov's solutions to the Monge–Ampère equation $\det \mathrm{D}^2 u = \nu$ with Dirichlet boundary conditions, where ν is a finitely supported measure. Global convergence for a damped Newton algorithm is established by Mirebeau [28] for a variant of Oliker and Prussner's discretization, but without convergence rates. Theorem 1.5 can be seen as an extension of the strategy applied by Mirebeau to optimal transport problems, which amounts to (a) replacing the Dirichlet boundary conditions with the second boundary value conditions from optimal transport, (b) replacing the Lebesgue measure by more general probability densities, and (c) changing the Monge–Ampère equation itself in order to deal with more general cost functions.

We also comment here that our result Theorem 5.1 answers a conjecture first raised by Gangbo and McCann in the case when the cost function satisfies the Ma–Trudinger–Wang condition. In [15, Example 1.6], a numerical approach to the semi-discrete optimal transport problem is suggested by taking what is equivalent to the negative gradient flow of the Kantorovich function defined in (1.3) above. There, Gangbo and McCann conjecture that this gradient flow should converge, and our result of uniform concavity of the Kantorovich functional confirms a quantitative strengthening of this conjecture, at least for costs, measures, and domains satisfying the assumptions of Theorem 5.1.

Finally, we note that the overall strategy for proving the convergence of Algorithm 1 (proving regularity then strict concavity of Φ) shares features of the one used in [9] to study the relationship between highly anisotropic semi-discrete quadratic optimal transport and Knothe rearrangement.

Outline. In Section 2, we establish the differentiability of Kantorovich's functional Φ, adapting arguments from [5]. In Sections 3 and 4, we prove the (uniform) second-differentiability of Kantorovich's functional when the cost function satisfies Loeper's (QC) condition. Section 5 is devoted to the proof of uniform concavity of Kantorovich's functional when the probability density satisfies a Poincaré–Wirtinger inequality (PW). In Section 6, we combine these intermediate results to prove the convergence of the damped Newton algorithm (Theorem 1.5), and we present a numerical illustration. Appendix A presents an explicit construction of a probability density with non-convex support over $\mathbb{R}^d$ which satisfies the assumptions of Theorem 1.5. Appendix B contains the details of the proof of the main theorem of Section 4.

2. Kantorovich's functional

The purpose of this section is to present the variational formulation introduced in [5] for the semi-discrete optimal transport problem, adapting the arguments presented [5] for the squared Euclidean cost in to cost functions satisfying (Reg′) and (Twist′), which are weaker than the conditions (Reg) and (Twist) presented in the introduction:

$$\forall y \in Y, \quad c(\cdot, y) \in \mathcal{C}^0(\Omega), \tag{Reg$'$}$$

$$\forall y \neq z \in Y,\ \forall t \in \mathbb{R}, \quad \mathcal{H}^d_g((c(\cdot, y) - c(\cdot, z))^{-1}(t)) = 0. \tag{Twist$'$}$$

Note that under (Twist$'$), the map $T_\varphi : X \to Y$ defined by (1.2) is uniquely defined $\mathcal{H}^d_g$-almost everywhere. Most of the results presented here are well known in the optimal transport literature; however, we include proofs for completeness.

Theorem 2.1. *Assume* (Reg$'$) *and* (Twist$'$), *and let ρ be a bounded probability density on X and $\nu = \sum_{y\in Y} \nu_y \delta_y$ a probability measure over Y. Then the functional Φ defined by* (1.3) *is concave, $\mathcal{C}^1$-smooth, and its gradient is given by* (1.4).

The proof of Theorem 2.1 relies on Propositions 2.2 and 2.3.

Proposition 2.2. *For any $\psi : Y \to \mathbb{R}$, the map T_ψ is an optimal transport map for the cost c between any probability density ρ on Ω and the push-forward measure $\nu := T_{\psi\#}\rho$.*

Proof. Assume that $\nu = S_\# \rho$ where S is a measurable map between X and Y. Then, by definition of T_ψ,

$$\forall x \in X, \quad c(x, T_\psi(x)) + \psi(T_\psi(x)) \leq c(x, S(x)) + \psi(S(x)).$$

Multiplying this inequality by ρ and integrating it over X gives

$$\int_X \big(c(x, T_\psi(x)) + \psi(T_\psi(x))\big)\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x) \leq \int_X \big(c(x, S(x)) + \psi(S(x))\big)\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x).$$

Since $\nu = S_\#\rho = T_{\psi\#}\rho$, the change of variable formula gives

$$\int_X \psi(S(x))\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x) = \int_Y \varphi(y)\, \mathrm{d}\nu = \int_X \psi(T_\psi(x))\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x).$$

Subtracting this equality from the inequality above shows that T_ψ is optimal:

$$\int_X c(x, T_\psi(x))\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x) \leq \int_X c(x, S(x))\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x). \qquad \square$$

Proposition 2.3. *Assume* (Twist$'$) *and* (Reg$'$). *Let ρ be a probability density over a compact subset X of Ω. Then the map $G : \mathbb{R}^Y \to \mathbb{R}^Y$ is continuous, where*

$$G(\psi) = (\rho(\mathrm{Lag}_y(\psi)))_{y\in Y}. \tag{2.1}$$

Lemma 2.4. *Let ρ be a probability density over a compact subset X of Ω, and let f in $\mathcal{C}^0(X)$ be such that $\rho(f^{-1}(t)) = 0$ for all $t \in \mathbb{R}$. Then the function $g : t \mapsto \rho(f^{-1}((-\infty, t]))$ is continuous.*

Proof. We consider the function $h(t) = \rho(f^{-1}((-\infty, t)))$. By hypothesis, $g(t) - h(t) = \rho(f^{-1}(t)) = 0$. Using Lebesgue's monotone convergence theorem one easily sees that g (resp. h) is right-continuous (resp. left-continuous). This concludes the proof. □

Proof of Proposition 2.3. Proving the continuity of G amounts to proving the continuity of the functions $G_y(\psi) := \rho(\mathrm{Lag}_y(\psi))$ for any y in Y. Fix y in Y and observe that by definition, $\mathrm{Lag}_y(\psi) = \bigcap_{z \neq y \in Y} H_z(\psi)$ where

$$H_z(\psi) := \{x \in X \mid c(x, y) + \psi(y) \leq c(x, z) + \psi(z)\}.$$

Denoting by $A \bigtriangleup B$ the symmetric difference of two sets, we have the inequalities

$$|G_y(\psi) - G_y(\varphi)| \leq \rho(\mathrm{Lag}_y(\psi) \bigtriangleup \mathrm{Lag}_y(\varphi)) \leq \sum_{z \in Y \setminus \{y\}} \rho(H_z(\psi) \bigtriangleup H_z(\varphi)). \tag{2.2}$$

Fix $z \neq y \in Y$, and denote $f = c(\cdot, y) - c(\cdot, z)$. Then

$$H_z(\psi) \bigtriangleup H_z(\varphi) \subseteq f^{-1}([\psi(z) - \psi(y), \varphi(z) - \varphi(y)]).$$

Here and below, we use the convention that $[a, b] = [\min\{a, b\}, \max\{a, b\}]$. By (Twist′) and Lemma 2.4 we know that $\lim_{\varphi \to \psi} \rho(H_z(\psi) \bigtriangleup H_z(\varphi)) = 0$, which by (2.2) concludes the proof. □

2.1. *Proof of Theorem 1.1*

We simultaneously show that the functional is concave and compute its gradient. For any function ψ on Y and any measurable map $T : X \to Y$, one has $\min_{y \in Y}(c(x, y) + \psi(y)) \leq c(y, T(y)) + \psi(T(y))$, which by integration gives

$$\Phi(\psi) \leq \int_X \big(c(x, T(x)) + \psi(T(x))\big)\rho(x)\, \mathrm{d}\mathcal{H}^d_g(x) - \sum_{y \in Y} \psi(y)\nu_y. \tag{2.3}$$

Moreover, equality holds when $T = T_\psi$. Taking another function φ on Y and setting $T = T_\varphi$ in (2.3) gives

$$\Phi(\psi) \leq \Phi(\varphi) + \langle G(\varphi) - \nu \mid \psi - \varphi \rangle,$$

where G is as in Proposition 2.3. This proves that the superdifferential $\partial^+\Phi(\varphi)$ contains $G(\varphi) - \nu$, thus establishing the concavity of Φ and its differentiability almost everywhere. It is known by [30, Theorem 25.6] that

$$\partial^+\Phi(\varphi) = \mathrm{conv}\left\{\lim_{n\to\infty} \nabla\Phi(\varphi_n) \;\middle|\; (\varphi_n) \in S\right\},$$

where conv denotes the convex envelope and S the set of sequences (φ_n) converging to φ such that Φ is differentiable at φ_n. By Proposition 2.3, the map G is continuous, meaning that we have constructed a continuous selection of the superdifferential of the concave function Φ:

$$\partial^+\Phi(\varphi) = \mathrm{conv}\left\{\lim_{n\to\infty} \nabla\Phi(\varphi_n)\right\} = \mathrm{conv}\left\{\lim_{n\to\infty} G(\varphi_n) - \nu\right\} = \{G(\varphi) - \nu\}.$$

This proves that Φ is $\mathcal{C}^1$, and that $\nabla\Phi(\varphi) = G(\varphi) - \nu$.

3. Local regularity in a c-exponential chart

The results presented in this section constitute an intermediate step in the proof of $\mathcal{C}^{2,\alpha}$ regularity of Kantorovich's functional. Let $\hat{X}$ be a compact, convex subset of $\mathbb{R}^d$ and $f_1, \dots, f_N$ be $\mathcal{C}^{1,1}$ functions on $\hat{X}$ which are quasi-convex, meaning that for any scalar $\lambda \in \mathbb{R}$ the closed sublevel sets $K_i(\lambda) := f_i^{-1}([-\infty, \lambda])$ are convex. Let $\hat{\rho}$ be a continuous probability density over $\hat{X}$. The purpose of this section is to give sufficient conditions for the regularity of the following function $\hat{G}$ near the origin of $\mathbb{R}^N$:

$$\hat{G} : \mathbb{R}^N \ni \boldsymbol{\lambda} \mapsto \int_{K(\boldsymbol{\lambda})} \hat{\rho}(x) \, \mathrm{d}\mathcal{H}^d(x), \tag{3.1}$$

where

$$K(\boldsymbol{\lambda}) := \bigcap_{i=1}^{N} K_i(\lambda_i) = \{x \in \hat{X} \mid \forall i \in \{1, \dots, N\}, \ f_i(x) \le \lambda_i\}.$$

3.1. Assumptions and statement of the theorem

We will impose two conditions on the functions $(f_i)_{1\le i\le N}$. As we will see in Section 4, both conditions are satisfied when these functions are constructed from a semi-discrete optimal transport transport problem whose cost function satisfies Loeper's condition (see Definition 1.1).

Non-degeneracy. The functions (f_i) satisfy the *non-degeneracy condition* if the norms of their gradients are bounded from below by a positive constant:

$$\varepsilon_{\mathrm{nd}} := \min_{1\le i\le N} \min_{\hat{X}} \|\nabla f_i\| > 0. \tag{ND}$$

This condition is necessary for the continuity of the map $\hat{G}$ even when $N = 1$.

Transversality. The boundary of the convex set $K(\boldsymbol{\lambda})$ can be decomposed into $N + 1$ *facets*, namely $(K(\boldsymbol{\lambda}) \cap \partial K_i(\lambda_i))_{1\le i\le N}$ and $K(\boldsymbol{\lambda}) \cap \partial \hat{X}$. The purpose of the transversality condition we consider is to ensure that the angle between adjacent facets is bounded from below by a positive constant when $\boldsymbol{\lambda}$ remains close to some fixed vector $\boldsymbol{\lambda}_0$.

Definition 3.1 (Normal cone). Let K be a convex compact set of $\mathbb{R}^d$. The *normal cone* to K at a point x in K is the set

$$\mathcal{N}_x K = \{v \in \mathbb{R}^d \mid \forall y \in K, \ \langle y - x \mid v\rangle \le 0\}, \tag{3.2}$$

and its elements are said to be *normal to K at x*.

Definition 3.2 (Transversality). The family (f_i) of functions satisfy the *transversality condition near $\boldsymbol{\lambda}_0$* if there exist positive constants $\varepsilon_{\mathrm{tr}}$ and $T_{\mathrm{tr}} \le 1$ such that for every $\boldsymbol{\lambda}$ in $\mathbb{R}^N$ satisfying $\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_\infty \le T_{\mathrm{tr}}$ for the usual ℓ^∞ norm on $\mathbb{R}^N$ and every point x in $\partial K(\boldsymbol{\lambda})$ one has:

$$\begin{aligned}&\text{if } \exists i \neq j \in \{1,\dots,N\},\ f_i(x)=\lambda_i \text{ and } f_j(x)=\lambda_j,\\ &\qquad\qquad \text{then } \left(\frac{\langle \nabla f_i(x) \mid \nabla f_j(x)\rangle}{\|\nabla f_i(x)\|\,\|\nabla f_j(x)\|}\right)^2 \leq 1-\varepsilon_{\mathrm{tr}}^2;\\ &\text{if } \exists i \in \{1,\dots,N\},\ f_i(x)=\lambda_i \text{ and } x\in\partial\hat{X},\\ &\qquad\qquad \text{then } \forall u \in \mathcal{N}_x\hat{X},\ \left(\frac{\langle u \mid \nabla f_j(x)\rangle}{\|u\|\,\|\nabla f_j(x)\|}\right)^2 \leq 1-\varepsilon_{\mathrm{tr}}^2.\end{aligned} \tag{T}$$

Note that if $\partial\hat{X}$ is smooth at x, then $\mathcal{N}_x\hat{X}$ is the ray spanned by the exterior normal to $\hat{X}$ at x.

Theorem 3.1. *Assume that the functions f_i satisfy the non-degeneracy condition (ND) and the transversality condition (T) near $\boldsymbol{\lambda}_0$. Let $\hat{\rho}$ be a $\mathcal{C}^{0,\alpha}$ probability density on $\hat{X}$. Then the map $\hat{G}$ defined in (3.1) is of class $\mathcal{C}^{1,\alpha}$ on the cube $Q := \boldsymbol{\lambda}_0 + [-T_{\mathrm{tr}}, T_{\mathrm{tr}}]^N$ and has partial derivatives given by*

$$\frac{\partial \hat{G}}{\partial \lambda_i}(\boldsymbol{\lambda}) = \int_{K(\boldsymbol{\lambda})\cap\partial K_i(\lambda_i)} \frac{\hat{\rho}(x)}{\|\nabla f_i(x)\|}\,\mathrm{d}\mathcal{H}^{d-1}(x). \tag{3.3}$$

In addition, the norm $\|\hat{G}\|_{\mathcal{C}^{1,\alpha}(Q)}$ is bounded by a constant depending only on $\varepsilon_{\mathrm{tr}}$, $\varepsilon_{\mathrm{nd}}$, $\|\hat{\rho}\|_{\mathcal{C}^{0,\alpha}(\hat{X})}$, on the diameter of $\hat{X}$ and on

$$C_M := \max_{1\leq i\leq N} \|\nabla f_i\|_\infty, \qquad C_L := \max_{1\leq i\leq N} \|\nabla f_i\|_{\mathrm{Lip}(\hat{X})}.$$

Note that the $\mathcal{C}^{1,\alpha}$ constant of $\hat{G}$ depends on the transversality constant $\varepsilon_{\mathrm{tr}}$ but does not depend on T_{tr}.

3.2. Sketch of proof

The correct expression for the partial derivatives of $\hat{G}$, given by (3.3), can easily be guessed by applying the coarea formula. The non-degeneracy condition then ensures that the denominator in this expression does not vanish. What is more delicate is to prove that these partial derivatives are α-Hölder, with a uniform estimate on the α-Hölder norm. A second application of the coarea formula on the manifold $f_i^{-1}(\lambda_i)$ suggests that for $j \neq i$ one should have

$$\left|\frac{\partial}{\partial \lambda_j}\int_{K(\boldsymbol{\lambda})\cap\partial K_i(\lambda_i)} \frac{\hat{\rho}(x)}{\|\nabla f_i(x)\|}\,\mathrm{d}\mathcal{H}^{d-1}(x)\right| \leq C\mathcal{H}^{d-2}\big(K(\boldsymbol{\lambda})\cap\partial K_i(\lambda_i)\cap\partial K_j(\lambda_j)\big)$$

under the assumption that the density $\hat{\rho}$ is $\mathcal{C}^1$ and the facet $K(\boldsymbol{\lambda})\cap\partial K_i(\boldsymbol{\lambda})$ does not intersect $\partial\hat{X}$. It will turn out that, thanks to the transversality hypothesis, the $\mathcal{H}^{d-2}$-measure of the union $\Sigma(\boldsymbol{\lambda})$ of these facets can be uniformly bounded:

$$\Sigma(\boldsymbol{\lambda}) = \bigcup_{1\leq i\leq N}\big(K(\boldsymbol{\lambda})\cap\partial\hat{X}\cap\partial K_i(\lambda_i)\big) \cup \bigcup_{1\leq i<j\leq N}\big(K(\boldsymbol{\lambda})\cap\partial K_i(\lambda_i)\cap\partial K_j(\lambda_j)\big).$$

Note also that equivalently, a point x belongs to the singular set $\Sigma(\boldsymbol{\lambda})$ if and only if it satisfies one of the assumptions in (T). In the next subsection, we prove an upper bound on $\mathcal{H}^{d-2}(\Sigma(\boldsymbol{\lambda}))$ (see Proposition 3.2). The proof of Theorem 3.1 follows from this upper bound and from several applications of the coarea formula. Since it is elementary but quite long, we have postponed the proof of the theorem itself to Appendix B.

3.3. Control on the $(d-2)$-Hausdorff measure of singular points

In this section, we prove that the transversality condition (T) and the quasi-convexity of the functions f_i imply a uniform upper bound on the $(d-2)$-Hausdorff measure of $\Sigma(\boldsymbol{\lambda})$.

Proposition 3.2. *Assuming the transversality condition* (T)*, there exists a constant depending only on d and* $\operatorname{diam}(\hat{X})$ *such that for* $\|\boldsymbol{\lambda}\|_\infty \leq T_{\mathrm{tr}}$,

$$\mathcal{H}^{d-2}(\Sigma(\boldsymbol{\lambda})) \leq C(d, \operatorname{diam}(\hat{X}))\frac{1}{\varepsilon_{\mathrm{tr}}}.$$

We will deduce this proposition from a general upper bound on the $(d-2)$-Hausdorff measure of the set of τ-singular points of a compact convex body. A more general and quantitative version of this bound can be found in [18]. Below we provide a straightforward and easy proof based on the notions of packing and covering numbers.

Proposition 3.3. *Let K be a convex, compact subset of $\mathbb{R}^d$ and $\tau > 0$. Then*

$$\mathcal{H}^{d-2}(\operatorname{Sing}(K,\tau)) \leq C(d, \operatorname{diam}(K))\frac{1}{\tau},$$

where $\operatorname{Sing}(K,\tau) := \{x \in \partial K \mid \exists u, v \in \mathcal{N}_x(K) \cap \mathbb{S}^{d-1},\ \langle u \mid v\rangle^2 \leq 1-\tau^2\}$.

Recall that the *covering number* $\operatorname{Cov}(K,\eta)$ of a subset $K \subseteq \mathbb{R}^d$ is the minimum number of Euclidean balls of radius η required to cover K. The *packing number* of a subset K is given by

$$\operatorname{Pack}(K,\eta) := \max\{\operatorname{Card}(X) \mid X \subseteq K \text{ and } \forall x \neq y \in X,\ \|x-y\| \geq \eta\}.$$

We will use the following comparisons between covering and packing numbers:

$$\operatorname{Cov}(K,\eta) \leq \operatorname{Pack}(K,\eta) \leq \operatorname{Cov}(K,\eta/2). \tag{3.4}$$

Proof of Proposition 3.3. The proof consists in comparing a lower bound and an upper bound of the packing number of the set

$$U := \{(x,n) \in \mathbb{R}^d \times \mathbb{S}^{d-1} \mid x \in \operatorname{Sing}(K,\tau) \text{ and } n \in \mathcal{N}_x(K)\}.$$

Step 1. We first calculate an upper bound on the covering number of the unit bundle $\mathcal{U}K := \{(x,n) \in \partial K \times \mathbb{S}^{d-1} \mid n \in \mathcal{N}_x K\}$. Given $r > 0$, we denote by K^r the set of points within distance r of K. By convexity, the projection map $\mathrm{p}_K : \mathbb{R}^d \to K$, mapping a point to its orthogonal projection on K, is well defined and 1-Lipschitz. We consider

$$\pi : \partial K^r \to \mathcal{U}(K), \qquad x \mapsto \left(\mathrm{p}_K(x), \frac{x - \mathrm{p}_K(x)}{\|x - \mathrm{p}_K(x)\|}\right).$$

The map π is surjective and has Lipschitz constant $L := \sqrt{1 + 4/r^2}$. We deduce an upper bound on the covering number of $\mathcal{U}K$ from the covering number of the level set ∂K^r:

$$\mathrm{Cov}(\mathcal{U}(K), \varepsilon) \leq \mathrm{Cov}(\partial K^r, \varepsilon/L).$$

Now, consider a sphere S with diameter $2\,\mathrm{diam}(K)$ that encloses the tubular neighborhood K^r with $r := \mathrm{diam}(K)$. The projection map p_{K^r} is 1-Lipschitz, and $\mathrm{p}_{K^r}(S) = \partial K^r$. Using the same argument as above, we have

$$\mathrm{Cov}(\partial K^r, \eta) \leq \mathrm{Cov}(S, \eta) \leq C(d) \cdot (\mathrm{diam}(K)/\eta)^{d-1}.$$

Combining these bounds with the inclusion $U \subseteq \mathcal{U}(K)$ gives

$$\mathrm{Cov}(U, \varepsilon) \leq \frac{C(d, \mathrm{diam}(K))}{\varepsilon^{d-1}}. \tag{3.5}$$

Step 2. We now establish a lower bound for $\mathrm{Pack}(U, 2\varepsilon)$. Let x be a τ-singular point and u, v be two unit vectors such that $\langle u \mid v\rangle^2 \leq 1 - \tau^2$. This implies that $\mathcal{N}_x K \cap \mathbb{S}^{d-1}$ contains a spherical geodesic segment of length at least $C \cdot \tau$, giving us a lower bound on the packing number of $\mathcal{N}_x K \cap \mathbb{S}^{d-1}$, namely $\mathrm{Pack}(\mathcal{N}_x K \cap \mathbb{S}^{d-1}, \eta) \geq C \cdot \tau/\eta$. Now, let X be a maximal set in the definition of the packing number $\mathrm{Pack}(\mathrm{Sing}(K, \tau), 2\varepsilon)$, and for every $x \in X$, let Y_x be a maximal set in the definition of the packing number $\mathrm{Pack}(\mathcal{N}_x(K) \cap \mathbb{S}^{d-1}, 2\varepsilon)$, so that $\mathrm{Card}(Y_x) \geq C \cdot \tau/\varepsilon$. Then the set $Z := \{(x, y) \mid x \in X,\ y \in Y_x\}$ is a 2ε-packing of U, and the cardinality of this set is bounded from below by $C \cdot \mathrm{Card}(X) \cdot \tau/\varepsilon$. This gives

$$\mathrm{Pack}(U, 2\varepsilon) \geq C \cdot \mathrm{Pack}(\mathrm{Sing}(K, \tau), 2\varepsilon) \cdot \tau/\varepsilon. \tag{3.6}$$

Step 3. Combining (3.5), (3.6) and the comparison between packing and covering numbers (3.4), we get

$$\mathrm{Pack}(\mathrm{Sing}(K, \tau), 2\varepsilon) \leq \frac{C(d, \mathrm{diam}(K))}{\tau \varepsilon^{d-2}}.$$

Using the comparison between packing and covering numbers, this means that we can cover $\mathrm{Sing}(K, \tau)$ with N_ε balls of radius ε such that $N_\varepsilon \leq C(d, \mathrm{diam}(K))/(\tau\varepsilon^{d-2})$. By definition of the Hausdorff measure, we have

$$\mathcal{H}^{d-2}(\mathrm{Sing}(K, \tau)) \leq \liminf_{\varepsilon \to 0} N_\varepsilon \varepsilon^{d-2} \leq C(d, \mathrm{diam}(K)) \frac{1}{\tau}. \qquad \square$$

Proof of Proposition 3.2. Given $\|\boldsymbol{\lambda}\|_\infty \leq T_{\mathrm{tr}}$, the transversality condition (T) implies

$$\forall x \in \Sigma(\boldsymbol{\lambda}),\ \exists u, v \in \mathcal{N}_x K(\boldsymbol{\lambda}), \qquad \left(\frac{\langle u \mid v\rangle}{\|u\|\,\|v\|}\right)^2 \leq 1 - \varepsilon_{\mathrm{tr}}^2,$$

where $\mathcal{N}_x K(\boldsymbol{\lambda})$ is the normal cone to the convex set $K(\boldsymbol{\lambda})$ at x (see (3.2)). This implies that $\Sigma(\boldsymbol{\lambda})$ is included in the set $\mathrm{Sing}(K(\boldsymbol{\lambda}), \varepsilon_{\mathrm{tr}})$ of τ-singular points with $\tau = \varepsilon_{\mathrm{tr}}$. The conclusion then follows from Proposition 3.3. $\square$

4. $\mathcal{C}^{2,\alpha}$ regularity of Kantorovich's functional

This section is devoted to the proof of the following regularity result. Recall that the conditions (Reg), (Twist), and (QC) are defined in the introduction on pages 2604, 2605, and 2608 respectively.

Theorem 4.1. *Assume* (Reg), (Twist)*, and* (QC). *Let X be a compact, c-convex subset of Ω and ρ in $\mathcal{P}^{\mathrm{ac}}(X) \cap \mathcal{C}^{0,\alpha}(X)$ for some α in* $(0, 1]$. *Then the Kantorovich functional Φ is uniformly $\mathcal{C}^{2,\alpha}$ on the set*

$$\mathcal{K}^{\varepsilon} := \{\psi : Y \to \mathbb{R} \mid \forall y \in Y,\ \rho(\mathrm{Lag}_y(\psi)) > \varepsilon\} \tag{4.1}$$

and its Hessian is given by (1.8). *In addition, the $\mathcal{C}^{2,\alpha}$ norm of the restriction of Φ to $\mathcal{K}^{\varepsilon}$ depends only on $\|\rho\|_{\infty}$, ε,* $\mathrm{diam}(X)$*, and the constants defined in Remark 4.1 below.*

For the remainder of the section, for any point y in Y, we will denote by $X_y = (\exp_y^c)^{-1}(X) \subseteq \mathbb{R}^d$ the inverse image of the domain X in the exponential chart at y. The set X_y is convex by c-concavity of X. We consider the functions

$$f_{z,y} : X_y \ni p \mapsto c(\exp_y^c(p), y) - c(\exp_y^c(p), z),$$

which are quasi-concave by (QC). The main difficulty in deducing Theorem 4.1 from Theorem 3.1 is to establish the quantitative transversality condition (T) introduced on p. 2615 for the family $(f_{z,y})_{z \in Y \setminus \{y\}}$.

Remark 4.1 (Constants). The $\mathcal{C}^{2,\alpha}$ norm of the restriction of Φ to $\mathcal{K}^{\varepsilon}$ explicitly depends on the following constants, whose finiteness (or positivity) follows from the compactness of the domain X, the finiteness of the set Y and the conditions (Reg), (Twist), and (QC):

$$\begin{aligned}
\varepsilon_{\mathrm{tw}} &:= \min_{x \in X} \min_{\substack{y,z \in Y\\ y \neq z}} \|\mathrm{D}_x c(x, y) - \mathrm{D}_x c(x, z)\|_g > 0,\\
C_{\nabla} &:= \max_{(x,y) \in X \times Y} \|\mathrm{D}_x c(x, y)\|_g < +\infty,\\
C_{\exp} &:= \max_{y \in Y} \max\{\|\exp_y^c\|_{\mathrm{Lip}(X_y)}, \|(\exp_y^c)^{-1}\|_{\mathrm{Lip}(X)}\} < +\infty,
\end{aligned} \tag{4.2}$$

where we recall that $X_y := \exp_y^{-1}(X)$. Our estimates will also rely on the following constants involving the differential of the exponential maps. As before, the tangent spaces $\mathrm{T}_x\Omega$ are endowed with the Riemannian metric g from Ω. We set

$$\begin{aligned}
C_{\mathrm{cond}} &:= \max_{y \in Y} \max_{p \in X_y} \mathrm{cond}(\mathrm{D}\exp_y^c|_p),\\
C_{\det} &:= \max_{y \in Y} \|\det(\mathrm{D}\exp_y^c)\|_{\mathrm{Lip}(X_y)},
\end{aligned}$$

where $\mathrm{cond}(A)$ is the condition number of a linear map A on a finite-dimensional normed space and $\det(A)$ is the determinant of A. The quantitative transversality estimates involve all the above constants in an explicit way (see (4.14)).

Remark 4.2. Even in the Euclidean case, one needs a lower bound on the volume of Laguerre cells in order to establish the second-differentiability of the functional Φ. Indeed, let $y_\pm = \pm 1$, $y_0 = 0$, and $Y = \{y_-, y_0, y_+\} \subset \mathbb{R}$. Consider the cost $c(x, y) = -xy$ and the density $\rho = 1$ on $X = [-1/2, 1/2]$. Let $\varphi_\tau \in \mathbb{R}^Y$ be defined by $\varphi(y_\pm) = 1/2$ and $\varphi(y_0) = \tau$. A simple calculation gives, for $\tau \geq 0$,

$$\frac{\partial \Phi}{\partial \mathbb{1}_{y_0}}(\psi_\tau) = \max(1 - 2\tau, 0),$$

which is not differentiable at $\tau = 1/2$, even though (Reg), (Twist), and (QC) are all satisfied.

Outline. In Section 4.1, we establish a part of the transversality condition using elementary properties of convex sets (Proposition 4.2). We establish in Section 4.2 a second transversality condition using additional assumptions and proceed in Section 4.3 to the proof of Theorem 4.1. In Section 4.4, we propose an alternative transversality estimate when Y is a sample subset of a target domain Ω' (Proposition 4.8).

4.1. Lower transversality estimates

Next, we undertake a series of proofs to obtain explicit constants in the transversality estimate (T), which depend on the choices of cost, domains, and dimension. Consider the Laguerre cell of a point y in Y in its own exponential chart, that is,

$$L_y(\psi) := (\exp_y^c)^{-1}(\mathrm{Lag}_y(\psi)) = \{p \in X_y \mid f_{z,y}(p) \leq \psi(z) - \psi(y)\}.$$

The set $L_y(\psi)$ is the intersection of sublevel sets of the functions $f_{z,y}$, and is therefore a convex subset of X_y by condition (QC). The first proposition establishes that two unit outer normals to L_y with the same basepoint cannot be near-opposite. Recall the definition of the normal cone from (3.2).

Proposition 4.2. *Assume that ψ lies in $\mathcal{K}^{\varepsilon/2}$ (see (4.1)). For any y in Y, any point p in $\partial L_y(\psi)$ and any unit normal vectors $v, w \in \mathcal{N}_p L_y(\psi)$ one has*

$$\langle v \mid w \rangle \geq -1 + \delta_0^2, \tag{4.3}$$

where $\delta_0 := \varepsilon/(2^{d-1}\|\rho\|_\infty C_{\exp}^{2d} \operatorname{diam}(X)^d) \leq 1$.

The proof of this proposition follows from a general lemma about convex sets. By convexity (QC), the set $L_y(\psi)$ is contained in an intersection of two half-spaces with outward normals v and w at p, giving an upper bound on its volume in term of its diameter and the angle between v and w (see Figure 2). On the other hand, we know that the volume of $L_y(\psi)$ is bounded from below by a constant depending on ε. Comparing these bounds will give us the one-sided estimate (4.3).

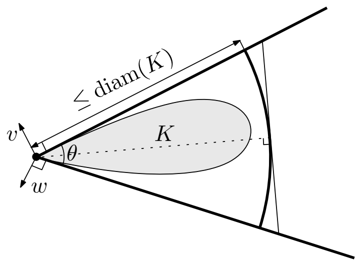

Fig. 2. Bound on the volume of a convex set K as a function of the angle between two normal vectors v, w at the same point and the diameter of K (see Lemma 4.3).

Lemma 4.3. *Let K be a bounded convex set in $\mathbb{R}^d$, let p be a boundary point of K and let v, w be two unit (outward) normal vectors to K at p. Then*

$$-1 + \delta_K^2 \leq \langle v \mid w \rangle \quad \textit{where} \quad \delta_K = \frac{\mathcal{H}^d(K)}{2^{d-2}\operatorname{diam}(K)^d} \leq 1.$$

Proof. The left-hand side of the inequality is non-positive, so the inequality needs only to be proven when $\langle v \mid w \rangle \leq 0$, which we assume from now on. Making a rotation of axes and a translation if necessary, we assume that p is the origin and the unit vectors span the first two coordinates of $\mathbb{R}^d$. Then, letting $H := \{p \mid \langle p \mid v \rangle \leq 0\}$, $H' := \{p \mid \langle p \mid w \rangle \leq 0\}$ and D be the two-dimensional disc centered at 0 of radius $\operatorname{diam}(K)$, one has

$$K \subseteq H \cap H' \cap (D \times [-\operatorname{diam}(K), \operatorname{diam}(K)]^{d-2}).$$

The intersection $H \cap H' \cap D$ is an angular sector of the disc D, whose angle is equal to $\theta := \pi - \arccos(\langle v \mid w \rangle)$ (see Figure 2). Therefore,

$$\begin{aligned}\mathcal{H}^d(K) &\leq \mathcal{H}^d\big(H \cap H' \cap (D \times [-\operatorname{diam}(K), \operatorname{diam}(K)]^{d-2})\big)\\ &\leq 2^{d-2}\operatorname{diam}(K)^d \tan(\theta/2).\end{aligned} \tag{4.4}$$

Using the expression of $\cos(\theta)$ in terms of $\tan(\theta/2)$ and recalling $\langle v \mid w \rangle \leq 0$ yields

$$\tan(\theta/2) = \sqrt{\frac{1 + \langle v \mid w \rangle}{1 - \langle v \mid w \rangle}} \leq \sqrt{1 + \langle v \mid w \rangle}. \tag{4.5}$$

The lemma follows directly from (4.4)–(4.5). □

Proof of Proposition 4.2. By definition of the bi-Lipschitz constant $C_{\exp}$,

$$\mathcal{H}^d(L_y(\psi)) \geq \varepsilon/(2C_{\exp}^d\|\rho\|_\infty) \quad \text{and} \quad \operatorname{diam}(L_y(\psi)) \leq C_{\exp}\operatorname{diam}(X).$$

Applying the above lemma to the two outward normals v, w at p, we get

$$\langle v \mid w \rangle + 1 \geq \frac{\mathcal{H}^d(L_y(\psi))^2}{4^{d-2}\operatorname{diam}(L_y(\psi))^{2d}} \geq \frac{\varepsilon^2}{4^{d-1}C_{\exp}^{4d}\|\rho\|_\infty^2\operatorname{diam}(X)^{2d}}. \qquad \square$$

We also record the following lemma for later use.

Lemma 4.4. *Let y be in Y and let p be a point of $L_y(\psi)$ such that for some $z \neq y$, $f_{z,y}(p) = \psi(z) - \psi(y)$. Then the point $p' := (\exp_z^c)^{-1}(\exp_y^c(p))$ belongs to $L_z(\psi)$ and the vector $\nabla f_{y,z}(p)$ lies in the normal cone $\mathcal{N}_{p'} L_z(\psi)$.*

Proof. We introduce the point $x = \exp_y^c(p)$. The hypothesis is equivalent to $c(x, y) + \psi(y) = c(x, z) + \psi(z)$. Since p belongs to $L_y(\psi)$, the point x belongs to $\mathrm{Lag}_y(\psi)$. Then, for any $z' \in Y$,

$$c(x, z) + \psi(z) = c(x, y) + \psi(y) \leq c(x, z') + \psi(z'),$$

thus establishing that $x \in \mathrm{Lag}_z(\psi)$ or equivalently $p' \in L_z(\psi)$. □

4.2. Upper transversality estimates

We now turn to the proof of the quantitative transversality estimates. We begin with a bound which involves the condition number of the differential of an exponential map (see Remark 4.1). The advantage of this bound is that we do not have to assume that the points in Y are sampled from a continuous domain. A second transversality estimate is presented in §4.4.

Notation. We introduce notation that will be used throughout this section. We fix a point y_0 in Y and an arbitrary ordering of the remaining points, so that $Y = \{y_0, y_1, \dots, y_N\}$. We define $\hat{X} := X_{y_0}$ and for every index $i \in \{1, \dots, N\}$ we put

$$f_i := f_{y_i, y_0} : \hat{X} \ni p \mapsto c(\exp_{y_0}^c(p), y_0) - c(\exp_{y_0}^c(p), y_i).$$

By the (Twist) condition, the functions $f_1, \dots, f_N$ satisfy the non-degeneracy condition (ND), and we have the following inequalities:

$$\varepsilon_{\mathrm{nd}} := \min_{i,j \neq 0} \min_{p \in X_{y_0}} \|\nabla f_i(p) - \nabla f_j(p)\| \geq C_{\exp}^{-1} \varepsilon_{\mathrm{tw}} > 0, \tag{4.6}$$

$$\sup_{i \neq 0} \sup_{p \in X_{y_0}} \|\nabla f_i(p)\| \leq C_{\exp} C_{\nabla}. \tag{4.7}$$

To any function $\psi : Y \to \mathbb{R}$ we associate the vector

$$\boldsymbol{\lambda}_\psi := (\psi(y_1) - \psi(y_0), \dots, \psi(y_N) - \psi(y_0)) \in \mathbb{R}^N. \tag{4.8}$$

We also consider the same family of convex set as in Section 3:

$$K(\boldsymbol{\lambda}) = \{p \in \hat{X} \mid \forall 1 \leq i \leq N,\ f_i(p) \leq \lambda_i\},$$

so that $K(\boldsymbol{\lambda}_\psi) = (\exp_{y_0}^c)^{-1}(\mathrm{Lag}_{y_0}(\psi))$.

Proposition 4.5. *Assume that $\boldsymbol{\lambda} := \boldsymbol{\lambda}_\psi$ where $\psi \in \mathcal{K}^{\varepsilon/2}$ and let $p \in K(\boldsymbol{\lambda})$.*

Case I: *If $f_i(p) = \lambda_i$ and $f_j(p) = \lambda_j$ for $i \neq j$ in $\{1, \dots, N\}$, then*

$$\left(\frac{\langle \nabla f_i(p) \mid \nabla f_j(p) \rangle}{\|\nabla f_i(p)\| \, \|\nabla f_j(p)\|} \right)^2 \leq 1 - \delta_1^2. \tag{4.9}$$

Case II: *If $p \in \partial \hat{X}$ and $f_i(p) = \lambda_i$ for some i in $\{1, \dots, N\}$, then*

$$\forall w \in \mathcal{N}_p \hat{X}, \quad \left(\frac{\langle \nabla f_i(p) \mid w \rangle}{\|\nabla f_i(p)\| \, \|w\|} \right)^2 \leq 1 - \delta_1^2. \tag{4.10}$$

In the above inequalities,

$$\delta_1 := \frac{\varepsilon_{\mathrm{nd}} \delta_0}{2 C_{\exp} C_\nabla C_{\mathrm{cond}}^2}.$$

By assumption, each Laguerre cell associated to ψ contains a mass of at least $\varepsilon/2$. This allows us to apply Proposition 4.2, ensuring that normal vectors to Laguerre cells in their exponential charts cannot be near-opposite. We denote $L_i := L_{y_i}(\psi) = (\exp^c_{y_i})^{-1}(\mathrm{Lag}_{y_i}(\psi))$ for brevity.

The proposition also relies on two simple lemmas. The first lemma shows the effect of a diffeomorphism on the normal cone to a convex set when its image is also convex.

Lemma 4.6. *Let $K \subset \mathbb{R}^d$ be a compact, convex set, let F be a $\mathcal{C}^1$ diffeomorphism from an open neighborhood of K to an open subset of $\mathbb{R}^d$, and assume that $F(K)$ is also a convex set. Then, for any point x in ∂K,*

$$\mathcal{N}_{F(x)}(F(K)) = [\mathrm{D}F^{-1}_{F(x)}]^*(\mathcal{N}_x K),$$

where A^* denotes the adjoint of A.

Proof. Consider $x \in \partial K$ and $v \in \mathcal{N}_x K$, and define $\varphi(z) := \langle F^{-1}(z) - x \mid v \rangle$. Since v is an outer normal to K at x, the restriction of φ to $F(K)$ is non-positive. Since $F(K)$ is convex, for any point $y \in K$ the set $F(K)$ contains the segment $[F(x), F(y)]$. Therefore

$$\begin{aligned} 0 &\geq \varphi((1-t)F(x) + tF(y)) \\ &\geq \varphi(F(x)) + t\langle \nabla\varphi(F(x)) \mid F(y) - F(x) \rangle - o(t) \\ &= t\langle [\mathrm{D}F^{-1}_{F(x)}]^*(v) \mid F(y) - F(x) \rangle - o(t), \end{aligned}$$

where we have used $\varphi(F(x)) = 0$ and $\nabla\varphi(F(x)) = [\mathrm{D}F^{-1}_{F(x)}]^*(v)$ to obtain the equality at the end. Dividing by t and taking the limit as t goes to zero, we see that

$$\forall y \in K, \quad \langle [\mathrm{D}F^{-1}_{F(x)}]^*(v) \mid F(y) - F(x) \rangle \leq 0,$$

thus showing that $[\mathrm{D}F^{-1}_{F(x)}]^*(v)$ belongs to the normal cone to $F(K)$ at $F(x)$. The converse inclusion follows from the symmetry of the problem. □

The second lemma compares the angle between two vectors and the angle between their images under a linear map, using the generalized Wiedlandt inequality (see [17, Section 3.4]). We identify $\mathbb{R}^d$ with its tangent and cotangent spaces through the Euclidean structure. We denote the adjoint of the derivative of the exponential map $\exp^c_y$ at a point p in X_y by

$$(\mathrm{D}\exp^c_{y_i})^*|_p : \mathrm{T}^*_{\exp^c_{y_i}(p)}\Omega \to \mathrm{T}^*_p\mathbb{R}^d \cong \mathbb{R}^d,$$

Lemma 4.7. *Let $y_k \neq y_\ell \in Y$, let $x \in X$ and set $p_k := (\exp^c_{y_k})^{-1}(x)$, $p_\ell := (\exp^c_{y_\ell})^{-1}(x)$ and*

$$A := (\mathrm{D}\exp^c_{y_k}|_{p_k})^* \circ [(\mathrm{D}\exp^c_{y_l}|_{p_l})^*]^{-1} : \mathrm{T}^*_{p_\ell}\mathbb{R}^d \to \mathrm{T}^*_{p_k}\mathbb{R}^d.$$

Then, for all v, w in $\mathbb{R}^d$,

$$C_{\mathrm{cond}}^{-4}\left(1 + \frac{\langle v \mid w\rangle}{\|v\|\,\|w\|}\right) \leq 1 + \frac{\langle Av \mid Aw\rangle}{\|Av\|\,\|Aw\|} \leq C_{\mathrm{cond}}^{4}\left(1 + \frac{\langle v \mid w\rangle}{\|v\|\,\|w\|}\right).$$

Proof. Indeed, let θ be the angle between v and w, and θ' the angle between Av and Aw, both in the interval $(0, \pi)$. Let $t := \tan(\theta/2)$ and $t' := \tan(\theta'/2)$. The generalized Wiedlandt inequality in [17, Section 3.4] asserts $(1/\mathrm{cond}(A))t \leq t' \leq \mathrm{cond}(A)t$. Expressing $\cos(\theta)$ in terms of $t = \tan(\theta/2)$, we obtain

$$1 + \cos(\theta') = 1 + \frac{1 - t'^2}{1 + t'^2} = \frac{2}{1 + t'^2} \leq \mathrm{cond}(A)^2(1 + \cos(\theta)).$$

We deduce the second inequality of the conclusion by using $\mathrm{cond}(A_2^*[A_1^*]^{-1}) \leq \mathrm{cond}(A_1)\,\mathrm{cond}(A_2)$ and the definition of the constant C_{cond}. For the first inequality, simply note that $\mathrm{cond}(A^{-1}) = \mathrm{cond}(A)$. □

Proof of Proposition 4.5, Case I. We let

$$V := \nabla f_i(p) = \nabla f_{y_i, y_0}(p), \qquad W := \nabla f_j(p) = \nabla f_{y_j, y_0}(p),$$
$$v := \frac{V}{\|V\|}, \qquad w := \frac{W}{\|W\|}.$$

Switching the indices i and j if necessary, we assume that $\|V\| \leq \|W\|$. The proof depends on the sign of $\langle W - V \mid V\rangle$ (see Remark 4.3 below for the significance of that sign). Assume first $\langle W - V \mid V\rangle \leq 0$, and let $\alpha_v := 1/\|V\|$ and $\alpha_w := 1/\|W\|$. Then

$$\begin{aligned} 1 - \langle v \mid w\rangle &= \tfrac{1}{2}\|v - w\|^2 = \tfrac{1}{2}\|\alpha_w(W - V) - (\alpha_v - \alpha_w)V\|^2 \\ &= \tfrac{1}{2}\alpha_w^2\|W - V\|^2 + \tfrac{1}{2}(\alpha_v - \alpha_w)^2\|V\|^2 - \alpha_w(\alpha_v - \alpha_w)\langle W - V \mid V\rangle. \end{aligned}$$

Using $\alpha_w \leq \alpha_v$ and $\|W - V\| \geq \varepsilon_{\mathrm{nd}}$ we end up with

$$\begin{aligned} 1 - \langle v \mid w\rangle^2 \geq 1 - \langle v \mid w\rangle &\geq \frac{1}{2}\alpha_w^2\|W - V\|^2 \\ &\geq \frac{1}{2}\,\frac{\varepsilon_{\mathrm{nd}}^2}{C_{\exp}^2 C_\nabla^2} \geq \frac{\varepsilon_{\mathrm{nd}}^2\delta_0^2}{4C_{\exp}^2 C_\nabla^2 C_{\mathrm{cond}}^4} = \delta_1^2, \end{aligned}$$

where we have used (4.6) and (4.7), $\delta_0 \leq 1$ and $C_{\mathrm{cond}} \geq 1$. This establishes the desired bound when $\langle v \mid w\rangle \in [0, 1]$. If $\langle v \mid w\rangle \in [-1, 0]$, we can apply Proposition 4.2 to show that $1 - \langle v \mid w\rangle^2 \geq 1 + \langle v \mid w\rangle \geq \delta_0^2 \geq \delta_1^2$, as desired.

Now suppose $\langle W - V \mid V \rangle \geq 0$. A slightly tedious computation gives

$$\langle v \mid w \rangle^2 = 1 - \frac{\|W - V\|^2}{\|W\|^2} + \frac{\langle W - V \mid v \rangle^2}{\|W\|^2} = 1 - \frac{\|W - V\|^2}{\|W\|^2}\left(1 - \frac{\langle W - V \mid v \rangle^2}{\|W - V\|^2}\right)$$
$$\leq 1 - \frac{\varepsilon_{\text{nd}}^2}{C_{\exp}^2 C_\nabla^2}\left(1 - \left\langle \frac{W - V}{\|W - V\|} \,\middle|\, v \right\rangle\right), \tag{4.11}$$

where we have used $\langle W - V \mid V \rangle \geq 0$ with (4.6) and (4.7) to get the last inequality. We will now apply Proposition 4.2 to the Laguerre cell L_i. By Lemma 4.4, the point $p_i := (\exp^c_{y_i})^{-1}(\exp^c_{y_0}(p)) \in X_{y_i}$ belongs to L_i and the vectors $V_i := \nabla f_{y_0, y_i}(p_i)$ and $W_i := \nabla f_{y_j, y_i}(p_i)$ are both normal to L_i at p_i. Proposition 4.2 then shows that the vectors V_i and W_i satisfy

$$-1 + \delta_0^2 \leq \frac{\langle V_i \mid W_i \rangle}{\|V_i\| \, \|W_i\|}. \tag{4.12}$$

We transfer this inequality to the exponential chart of the original point y_0 using the linear map

$$A := (\mathrm{D}\exp^c_{y_0}|_p)^* \circ [(\mathrm{D}\exp^c_{y_i}|_{p_i})^*]^{-1}.$$

First, note that $W - V = AW_i$ and $V = -AV_i$. Applying the generalized Wiedlandt inequality (Lemma 4.7) and (4.12) we have

$$1 - \frac{\langle W - V \mid v \rangle}{\|W - V\|} = 1 + \frac{\langle AW_i \mid AV_i \rangle}{\|AW_i\| \, \|AV_i\|} \geq C_{\text{cond}}^{-4}\left(1 + \frac{\langle V_i \mid W_i \rangle}{\|V_i\| \, \|W_i\|}\right)$$
$$\geq C_{\text{cond}}^{-4}\delta_0^2 \geq \delta_1^2. \tag{4.13}$$

Combining this inequality with (4.11) we obtain (4.9) in this case as well. □

Proof of Proposition 4.5, Case II. Consider $V := \nabla f_i(p)$ and let W be any vector in the normal cone $\mathcal{N}_p \hat{X}$. When $\langle V \mid W \rangle \leq 0$, the inequality directly follows from Proposition 4.2, ensuring that normal vectors cannot be near-opposite. We now assume $\langle V \mid W \rangle \geq 0$ and we will apply Proposition 4.2 to the Laguerre cell of y_i and transfer the result to the exponential chart of the point y_0. Let $p_i = (\exp^c_{y_i})^{-1}(\exp^c_y(p))$. Then, by Lemma 4.4, p_i belongs to L_i and $V_i := \nabla f_{y_0, y_i}(p_i)$ is a normal vector to L_i at p_i. We define a second normal vector by considering

$$A := (\mathrm{D}\exp^c_{y_0}|_{p_0})^* \circ [(\mathrm{D}\exp^c_{y_i}|_{p_i})^*]^{-1}$$

and by setting $W_i := A^{-1}W \in \mathrm{T}^*_{p_i}\mathbb{R}^d$. By Lemma 4.6, the vector W_i belongs to the normal cone to X_{y_i} at p_i. Moreover, since L_i is contained in X_{y_i} and both sets contain p_i, we have $\mathcal{N}_{p_i} X_{y_i} \subseteq \mathcal{N}_{p_i} L_i$, thus ensuring that W_i also belongs to the normal cone to L_i at p_i. Then, by Proposition 4.2 again,

$$\frac{\langle V_i \mid W_i \rangle}{\|V_i\| \, \|W_i\|} \geq -1 + \delta_0^2.$$

As before, we transfer this inequality to the exponential chart of the original point y using the linear map A. We have $V = \nabla f_i(p) = -AV_i$, and by construction $W = AW_i$. We get the desired inequality by applying Lemma 4.7:

$$1 - \frac{\langle V \mid W\rangle}{\|V\| \, \|W\|} = 1 + \frac{\langle AV_i \mid AW_i\rangle}{\|AV_i\| \, \|AW_i\|} \geq C_{\mathrm{cond}}^{-4}\left(1 + \frac{\langle V_i \mid W_i\rangle}{\|V_i\| \, \|W_i\|}\right)$$
$$\geq C_{\mathrm{cond}}^{-4}\delta_0^2 \geq \delta_1^2,$$

and by recalling that $\langle V \mid W\rangle \geq 0$. □

4.3. *Proof of Theorem 4.1*

By Theorem 1.1, the second-differentiability of Kantorovich's functional Φ will follow from the differentiability of the function

$$G_{y_0}(\psi) := \int_{\mathrm{Lag}_{y_0}(\psi)} \rho(x) \,\mathrm{d}\mathcal{H}_g^d(x) = \int_{L_y(\psi)} \hat{\rho}(p) \,\mathrm{d}p,$$

where we have used the change-of-variable formula with $x = \exp^c_{y_0}(p)$, so that $\hat{\rho}$ is the density of the push-forward measure $(\exp^c_{y_0})^{-1}_{\#}(\rho\mathcal{H}^d_g)$ with respect to the Lebesgue measure. We recall that

$$K(\boldsymbol{\lambda}_\psi) = (\exp^c_{y_0})^{-1}(\mathrm{Lag}_{y_0}(\psi)),$$

so that $G_{y_0}(\psi) = \hat{G}(\boldsymbol{\lambda}_\psi)$ (as defined in (3.1)). The differentiability of $\hat{G}$ will be proven using Theorem 3.1 from the previous section.

Let us fix a function ψ_0 in $\mathcal{K}^\varepsilon$ and recall that $\boldsymbol{\lambda}_0 := \boldsymbol{\lambda}_{\psi_0}$. By Proposition 2.3 there exists a positive constant T_{tr} such that every function ψ on Y satisfying $\|\psi - \psi_0\|_\infty \leq T_{\mathrm{tr}}$ belongs to $\mathcal{K}^{\varepsilon/2}$. Then, by Proposition 4.5, we see that the functions f_i satisfy the transversality condition (T) on the cube $\boldsymbol{\lambda}_0 + [-T_{\mathrm{tr}}, T_{\mathrm{tr}}]^N$ with constant

$$\varepsilon_{\mathrm{tr}} = \delta_1 = \frac{\varepsilon_{\mathrm{nd}}\delta_0}{2C_{\exp}C_\nabla C^2_{\mathrm{cond}}}, \tag{4.14}$$

where we recall that $\delta_0 = \varepsilon/(2^d\|\rho\|_\infty C^{2d}_{\exp}\,\mathrm{diam}(X)^d)$. Note also that since ρ is α-Hölder and since the exponential map is $\mathcal{C}^{1,1}$, the probability density $\hat{\rho}$ is also α-Hölder with constant

$$\|\hat{\rho}\|_{\mathcal{C}^{0,\alpha}(\hat{X})} \leq C(\|\rho\|_{\mathcal{C}^{0,\alpha}}, C_{\det}). \tag{4.15}$$

We can now apply Theorem 3.1. It ensures that the function $\hat{G}$ is of class $\mathcal{C}^{1,\alpha}$ on the cube $\boldsymbol{\lambda}_0 + [-T_{\mathrm{tr}}, T_{\mathrm{tr}}]^N$, so that $\partial\Phi/\partial\mathbb{1}_{y_0}$ is $\mathcal{C}^{1,\alpha}$ on a neighborhood of ψ_0. Since this holds for any point $y_0 \in Y$ and any function ψ_0 in $\mathcal{K}^\varepsilon$, we have established the $\mathcal{C}^{2,\alpha}$ regularity of Φ on $\mathcal{K}^\varepsilon$. The claimed dependency of $\|\Phi\|_{\mathcal{C}^{2,\alpha}(\mathcal{K}^\varepsilon)}$ follows from (4.14)–(4.15) and from Theorem 3.1.

Our goal is now to deduce the formula for the gradient of G given in Theorem 4.1 (equation (1.8)) from the formula for the gradient of $\hat{G}$ given in Theorem 3.1 (equation (3.3)).

This is done by looking more closely at the change of variable induced by the exponential map $F := \exp^c_{y_0} : \Omega \to \mathbb{R}^d$. For ease of notation we let $h := c(\cdot, y_0) - c(\cdot, y_i)\lambda = f_i \circ F^{-1}$. By the definition of push-forward, for any bounded measurable function χ on Ω we have

$$\int_{\hat{\Omega}} \chi(F(p))\rho(p)\,\mathrm{d}\mathcal{H}^d(p) = \int_{\Omega} \chi(x)\hat{\rho}(x)\,\mathrm{d}\mathcal{H}^d_g(x).$$

Multiplying χ by the characteristic function of $h^{-1}([t, s])$ gives

$$\int_{f_i^{-1}([t,s])} \chi(F(p))\hat{\rho}(p)\,\mathrm{d}\mathcal{H}^d(p) = \int_{h^{-1}([t,s])} \chi(x)\rho(x)\,\mathrm{d}\mathcal{H}^d_g(x).$$

Applying the coarea formula on both sides, we get

$$\int_t^s \int_{f_i^{-1}(r)} \frac{\chi(F(p))\hat{\rho}(p)}{\|\nabla f_i(p)\|}\,\mathrm{d}\mathcal{H}^{d-1}(p)\,\mathrm{d}r = \int_t^s \int_{h^{-1}(r)} \frac{\chi(x)\rho(x)}{\|\nabla h(x)\|_g}\,\mathrm{d}\mathcal{H}^{d-1}_g(x)\,\mathrm{d}r. \quad (4.16)$$

Using the $\mathcal{C}^{1,1}$ smoothness of the functions f_i and the (Twist) condition, we can see that for any χ in $\mathcal{C}^0_c(\Omega)$, the two inner integrals

$$r \mapsto \int_{f_i^{-1}(r)} \frac{\chi(F(p))\hat{\rho}(p)}{\|\nabla f_i(p)\|}\,\mathrm{d}\mathcal{H}^{d-1}(p) \quad \text{and} \quad r \mapsto \int_{h^{-1}(r)} \frac{\chi(x)\rho(x)}{\|\nabla h(x)\|_g}\,\mathrm{d}\mathcal{H}^{d-1}_g(x)$$

depend continuously on r. Using the continuity of these functions in r, equation (4.16) and the Fundamental Theorem of Calculus, we find that for any function χ in $\mathcal{C}^0_c(\Omega)$ and any r in $\mathbb{R}$,

$$\int_{f_i^{-1}(r)} \frac{\chi(F(p))\hat{\rho}(p)}{\|\nabla f_i(p)\|}\,\mathrm{d}\mathcal{H}^{d-1}(p) = \int_{h^{-1}(r)} \frac{\chi(x)\rho(x)}{\|\nabla h(x)\|_g}\,\mathrm{d}\mathcal{H}^{d-1}_g(x).$$

By Tietze's extension theorem, (Twist), and (Reg), the level set $S := h^{-1}(r)$ is a $\mathcal{C}^{1,1}$ hypersurface of Ω. Thus every function in $\mathcal{C}^0_c(S)$ can be extended to a function in $\mathcal{C}^0_c(\Omega)$. The previous equality therefore holds for any χ in $\mathcal{C}^0_c(S)$, and by density, also for any function χ in $L^1(S)$. Applying this with χ equal to the indicator function of the interface between the Laguerre cell of y_0 and the cell of y_i, we get the desired formula for the partial derivatives:

$$\begin{aligned}\frac{\partial G_i}{\partial \psi_i}(\psi) = \frac{\partial \hat{G}}{\partial \lambda_i}(\boldsymbol{\lambda}_\psi) &= \int_{L_{y_0}(\psi)\cap f_i^{-1}(\psi(y_i)-\psi(y_0))} \frac{\hat{\rho}(p)}{\|\nabla f_i(p)\|}\,\mathrm{d}\mathcal{H}^{d-1}(p)\\ &= \int_{\mathrm{Lag}_{y_0}(\psi)\cap \mathrm{Lag}_{y_i}(\psi)} \frac{\rho(x)}{\|\mathrm{D}_y c(x, y_0) - \mathrm{D}_y c(x, y_i)\|_g}\,\mathrm{d}\mathcal{H}^{d-1}_g(x).\end{aligned}$$

4.4. Alternative upper transversality estimates

Finally, we state an alternative upper transversality estimate, under the assumption that the points in Y are sampled from some *target domain* Λ, along with some convexity conditions. Specifically, let Λ be a bounded, open subset in some Riemannian manifold, with $Y \subset \Lambda$. We then assume that for any $x' \in \Omega^{\mathrm{cl}}$, the mapping

$$y \mapsto -\mathrm{D}_x c(x', y)$$

is a diffeomorphism onto its range, and we denote the inverse by $\exp^c_{x'}$. We will also assume that $(\exp^c_x)^{-1}(\Lambda)$ is convex for all $x \in \Omega$, and finally that for any $x, x' \in \Omega$, $q_0, q_1 \in (\exp^c_{x'})^{-1}(\Lambda)$, and $t \in [0, 1]$,

$$\begin{aligned} &-c(x, \exp^c_{x'}((1-t)q_0 + tq_1)) + c(x', \exp^c_{x'}((1-t)q_0 + tq_1))) \\ &\quad \le \max\{-c(x, \exp^c_{x'}(q_0)) + c(x', \exp^c_{x'}(q_0)), -c(x, \exp^c_{x'}(q_1)) + c(x', \exp^c_{x'}(q_1))\}. \end{aligned} \tag{4.17}$$

Note that this last inequality is nothing but quasi-concavity of $c(x', \cdot) - c(x, \cdot)$ in the global coordinate chart of Λ defined by $\exp^c_{x'}$. For more on these conditions, see Remark 4.4 below.

Proposition 4.8 can be applied to provide an alternative bound in the transversality condition (T) when the point $p_0 \in \partial K(\boldsymbol{\lambda})$ is in the interior of X (so in particular, when dealing with Laguerre cells that do not intersect ∂X). The advantage of this bound is that it does not require knowledge of the condition number C_{cond}.

Recall that we have fixed some point $y_0 \in Y = \{y_1, \ldots, y_N\}$ and for any index $i \in \{1, \ldots, N\}$ we use the notation

$$f_i(p) = f_{y_i, y_0}(p) = c(\exp^c_{y_0} p, y_0) - c(\exp^c_{y_0} p, y_i).$$

We also redefine the constants C_∇ and $C_{\exp}$ so that in their definitions, the maximum of y ranges over the domain Λ instead of just Y.

Proposition 4.8. *Suppose*

$$\|\boldsymbol{\lambda}\| < T_{\mathrm{tr}} < \frac{\varepsilon \varepsilon_{\mathrm{nd}}}{8 C_{\exp}^{2d-1} \|\rho\|_\infty \mathcal{H}^{d-1}_g(\partial X)},$$

and $p_0 \in K(\boldsymbol{\lambda})$ *with* $f_i(p_0) = \lambda_i$ *and* $f_j(p_0) = \lambda_j$ *for some* $i \neq j$ *in* $\{1, \ldots, N\}$. *Then*

$$\left(\frac{\langle \nabla f_i(p_0) \mid \nabla f_j(p_0) \rangle}{\|\nabla f_i(p_0)\| \, \|\nabla f_j(p_0)\|} \right)^2 \le (1 - \delta_2)^2, \tag{4.18}$$

where

$$\delta_2 := \left(\frac{\varepsilon \varepsilon_{\mathrm{nd}}}{4\sqrt{2}\, C_\nabla^2 C_{\exp}^{2d+2} \|\rho\|_\infty \mathcal{H}^{d-1}_g(\partial X)} \right)^2.$$

Remark 4.3. Before embarking on the proof of this "continuous" upper transversality estimate, we compare some key features of its proof with that of the "discrete" upper transversality estimate, Proposition 4.5. By considering the case when the two vectors $\nabla f_i(p_0)$ and $\nabla f_j(p_0)$ are collinear, we can see that both proofs rely on the same core idea. In this case, $\nabla f_i(p_0)$ and $\nabla f_j(p_0)$ are outward normal vectors (in coordinates induced by $\exp^c_{y_0}$, see Remark 4.4) to the sublevel sets $\{-c(\cdot, y_0) + \psi(y_0) \leq -c(\cdot, y_i) + \psi(y_i)\}$ and $\{-c(\cdot, y_0) + \psi(y_0) \leq -c(\cdot, y_j) + \psi(y_j)\}$ respectively at p_0 which lies on the intersection of their boundaries. Since these sets are convex in the associated coordinates, this will cause the Laguerre cell associated to y_i to be trapped in a lower-dimensional set, giving it zero mass, which is a contradiction. The difference between the two proofs lies in quantifying this estimate. In the discrete version of the estimate we do one of two things depending on the sign of the inner product $\langle W - V \mid V\rangle$ (see proof of Proposition 4.5). When the inner product $\langle W - V, V\rangle$ is negative, since $\|W\| \geq \|V\|$ and $\|W - V\|$ has a positive lower bound (by the condition (ND)), it can be seen that there is a cone whose axis is in the direction of V and whose opening angle can be estimated from below, and the vector W points to the outside of this cone. In the other case $W - V$ and $-V$ are, respectively, outward normal vectors to the sublevel sets $\{-c(\cdot, y_i) + \psi(y_i) \leq -c(\cdot, y_j) + \psi(y_j)\}$ and $\{-c(\cdot, y_i) + \psi(y_i) \leq -c(\cdot, y_0) + \psi(y_0)\}$, viewed in coordinates given by $\exp^c_{y_i}$. Thus the lower transversality estimate (Proposition 4.2) can be applied to obtain a quantitative bound, but at the price of involving the condition number since we have made a change of coordinates. In the continuous version, there is no change of coordinates, instead we make a rotation to align $\nabla f_j(p_0)$ and $\nabla f_i(p_0)$, then estimate the error induced by this rotation using (4.17), in a vein similar to calculations from [16, Remark 2.5, Proof of Lemma 4.7].

Proof of Proposition 4.8. Let us again write

$$V := \nabla f_i(p_0), \qquad W := \nabla f_j(p_0), \qquad v := \frac{V}{\|V\|}, \qquad w := \frac{W}{\|W\|},$$

and assume $\varepsilon_{\mathrm{nd}} < \|V\| \leq \|W\|$ and $\langle v \mid w\rangle > 0$. Let us also define

$$x_0 := \exp^c_{y_0}(p_0), \qquad q_0 := -\mathrm{D}_x c(x_0, y_0), \qquad q_1 := -\mathrm{D}_x c(x_0, y_j).$$

A quick calculation yields

$$\begin{aligned} q_0 &= [(\mathrm{D}\exp^c_{y_0}|_{p_0})^*]^{-1}(-\nabla_p c(\exp^c_{y_0}(p), y_0)|_{p=p_0}),\\ q_1 &= [(\mathrm{D}\exp^c_{y_0}|_{p_0})^*]^{-1}(W) + q_0. \end{aligned}$$

Now we define

$$q' := [(\mathrm{D}\exp^c_{y_0}|_{p_0})^*]^{-1}(\|V\|w) + q_0;$$

since $\|V\| \leq \|W\|$, the above calculation implies that q' lies on the line segment between q_0 and q_1; since $(\exp^c_{x_0})^{-1}(\Lambda)$ is convex, we have $q' \in (\exp^c_{x_0})^{-1}(\Lambda)$ as well.

Thus we can define

$$\begin{aligned} y_i' &:= \exp^c_{x_0}(q'),\\ \tilde{f}_i(p) &:= -c(\exp^c_{y_0}(p), y_i') + c(\exp^c_{y_0}(p), y_0) + c(x_0, y_i') - c(x_0, y_0) + \lambda_i, \end{aligned}$$

and by (4.17) applied with the choices $x := \exp^c_{y_0}(p)$, $x' := x_0$, $t := \|V\|/\|W\|$, and q_0, q_1 as defined above, we will obtain, for all $p \in (\exp^c_{y_0})^{-1}(\Omega)$,

$$\tilde{f}_i(p) - \lambda_i \leq \max\{0, f_j(p) - \lambda_j\}, \tag{4.19}$$

while another quick calculation yields

$$y_i = \exp^c_{x_0}([(\mathrm{D}\exp^c_{y_0}|_{p_0})^*]^{-1}(V) + q_0).$$

Now note that

$$\begin{aligned}
&|-c(\exp^c_{y_0}(p), y_i') + c(\exp^c_{y_0}(p), y_i)| \\
&\leq \sup_{(x,q)\in\Omega\times(\exp^c_{x_0})^{-1}(\Lambda)} \left\|(\mathrm{D}\exp^c_{x_0}|_q)^*(-\mathrm{D}_y c(x, \exp^c_{x_0}(q)))\right\| \\
&\qquad\qquad \cdot \left\|[(\mathrm{D}\exp^c_{y_0}|_{p_0})^*]^{-1}(\|V\|w - V)\right\| \\
&\leq C_\nabla C^2_{\exp} \left\|\|V\|w - V\right\|,
\end{aligned}$$

where we have used the fact that if $y = \exp^c_{x_0}(q)$, then $(\mathrm{D}\exp^c_{x_0}|_q)^* = \mathrm{D}\exp^c_y|_{-\mathrm{D}_x c(x_0,y)}$. As a result we obtain

$$\begin{aligned}
|\tilde{f}_i(p) - f_i(p)|^2 &= |-c(\exp^c_{y_0}(p), y_i') + c(\exp^c_{y_0}(p), y_i)|^2 \\
&\leq (C_\nabla C^2_{\exp})^2\left\|V - \|V\|w\right\|^2 = 2(C_\nabla C^2_{\exp})^2\|V\|^2(1 - \langle v \mid w\rangle) \\
&\leq 2(C_\nabla C^2_{\exp})^2(C_{\exp}C_\nabla)^2(1 - \langle v \mid w\rangle).
\end{aligned}$$

Combining this with (4.19) we then have, for any $p \in (\exp^c_{y_0})^{-1}(\Omega)$,

$$f_i(p) - \lambda_i \leq \max\{0, f_j(p) - \lambda_j\} + \sqrt{2}\, C^3_{\exp} C^2_\nabla \sqrt{1 - \langle v \mid w\rangle},$$

or after rearranging and using $\|\boldsymbol{\lambda}\| < T_{\mathrm{tr}}$,

$$1 - \langle v \mid w\rangle \geq \sup_{p\in(\exp^c_{y_0})^{-1}(\Omega)} \frac{(f_i(p) - \max\{0, f_j(p)\} - 2T_{\mathrm{tr}})^2}{2(C^3_{\exp}C^2_\nabla)^2}. \tag{4.20}$$

We now make the following observation. Let us write $X_i := (\exp^c_{y_i})^{-1}(X)$. Then for any $t, s > 0$, we can estimate the volume of $X_i \cap \{f_{y_0,y_i} \leq -t\} \cap \{f_{y_j,y_i} \leq -s\}$ from below by

$$\begin{aligned}
&\mathcal{H}^d(X_i \cap \{f_{y_0,y_i} \leq -t\} \cap \{f_{y_j,y_i} \leq -s\}) \\
&\qquad \geq \mathcal{H}^d(X_i \cap \{f_{y_0,y_i} \leq 0\} \cap \{f_{y_j,y_i} \leq 0\}) - \mathcal{H}^d(X_i \cap \{-t < f_{y_0,y_i} \leq 0\}) \\
&\qquad\quad - \mathcal{H}^d(X_i \cap \{-s < f_{y_j,y_i} \leq 0\}).
\end{aligned}$$

Using $L_i \subset \{f_{y_0,y_i} \leq 0\} \cap \{f_{y_j,y_i} \leq 0\}$, we can bound the first term from below as

$$\mathcal{H}^d(X_i \cap \{f_{y_0,y_i} \leq 0\} \cap \{f_{y_j,y_i} \leq 0\}) \geq \frac{\varepsilon}{C^d_{\exp}\|\rho\|_\infty}.$$

For the second term, by the coarea formula we can write

$$\mathcal{H}^d(X_i \cap \{-t < f_{y_0,y_i} \leq 0\}) \leq \int_{-t}^{0} \int_{X_i \cap \{f_{y_0,y_i}=z\}} \frac{1}{\|\nabla f_{y_0,y_i}(p)\|} \, d\mathcal{H}^{d-1}(p) \, dz$$
$$\leq \frac{t\mathcal{H}^{d-1}(\partial X_i)}{\varepsilon_{\mathrm{nd}}} \leq \frac{tC_{\exp}^{d-1}\mathcal{H}_g^{d-1}(\partial X)}{\varepsilon_{\mathrm{nd}}},$$

where to obtain the second line we have again used the fact that for every $z \in \mathbb{R}$ the set $X_i \cap \{f_{y_0,y_i} = z\}$ is contained in the boundary of a convex subset of X_i, in conjunction with [21, Remark 5.2]. By a similar bound on the third term, we see that as long as

$$\max\{t, s\} < \frac{\varepsilon\varepsilon_{\mathrm{nd}}}{2C_{\exp}^{2d-1}\|\rho\|_\infty \mathcal{H}_g^{d-1}(\partial X)}$$

we have

$$\mathcal{H}^d(X_i \cap \{f_{y_0,y_i} \leq -t\} \cap \{f_{y_j,y_i} \leq -s\}) > 0,$$

thus in particular (by continuity of f_{y_0,y_i} and f_{y_j,y_i}) there must exist a point $p'_c \in X_i$ for which

$$\max\{f_{y_0,y_i}(p'_c), \ f_{y_j,y_i}(p'_c)\} \leq -\frac{\varepsilon\varepsilon_{\mathrm{nd}}}{2C_{\exp}^{2d-1}\|\rho\|_\infty \mathcal{H}_g^{d-1}(\partial X)}.$$

Translating this back into coordinates in $(\exp_0^c)^{-1}(X)$ and in terms of f_i, f_j, we see there exists a point $p_c \in (\exp_0^c)^{-1}(X)$ for which

$$f_i(p_c) - \max\{0, f_j(p_c)\} \geq \frac{\varepsilon\varepsilon_{\mathrm{nd}}}{2C_{\exp}^{2d-1}\|\rho\|_\infty \mathcal{H}_g^{d-1}(\partial X)}.$$

Thus if $T_{\mathrm{tr}} \leq \frac{\varepsilon\varepsilon_{\mathrm{nd}}}{8C_{\exp}^{2d-1}\|\rho\|_\infty \mathcal{H}_g^{d-1}(\partial X)}$, then combining this with (4.20) we will obtain the bound (4.18) as desired. □

Remark 4.4. Under a set of standard conditions, we can obtain both (QC) and (4.17).

Let Ω and Λ be bounded and smooth domains in d-dimensional Riemannian manifolds and take a cost $c \in \mathcal{C}^4(\overline{\Omega} \times \overline{\Lambda})$. Also assume:

- c satisfies the (Twist) condition: for every $x \in \Omega$, the map $\Lambda \ni y \mapsto -\mathrm{D}_x c(x, y)$ is a diffeomorphism onto its image $\Lambda_x := -\mathrm{D}_x c(x, \Lambda) \subset T_x^*\Omega$ and we define the c-exponential map $\exp_x^c : \Lambda_x \to \Lambda$ by $\exp_x^c = (-\mathrm{D}_x c(x, \cdot))^{-1}$.
- the cost $c^*(x, y) := c(y, x)$ satisfies the (Twist) condition: for every $y \in \Lambda$, we can define the c^*-exponential map $\exp_y^c : \Omega_y \to \Omega$ on the set $\Omega_y := -\mathrm{D}_y c(\Omega, y) \subset T_y^*\Lambda$ by $\exp_y^c = (-\mathrm{D}_y c(\cdot, y))^{-1}$.
- $(\exp_x^c)^{-1}(\Lambda)$ is convex for each $x \in \Omega$.
- $(\exp_y^{c^*})^{-1}(\Omega)$ is convex for each $y \in \Lambda$.
- $\det \mathrm{D}_{xy}^2 c(x, y) \neq 0$ for all $(x, y) \in \overline{\Omega} \times \overline{\Lambda}$.

- For any $(x, y) \in \overline{\Omega} \times \overline{\Lambda}$ and $\eta \in T_x^*\Omega$, $V \in T_x\Omega$ with $\eta(V) = 0$,

$$-(c_{ij,pq} - c_{ij,r}c^{r,s}c_{s,pq})c^{p,k}c^{q,l}V^iV^j\eta_k\eta_l \geq 0; \tag{A3w}$$

here indices before a comma are derivatives on Ω and after a comma on Λ, for fixed coordinate systems, and a pair of raised indices denotes the inverse of a matrix. This last condition (A3w) originates (in a stronger version) in [26] related to regularity of optimal transport. [23, Theorem 3.2] in the Euclidean case and [20, Theorem 4.10] in the general manifold case show the above conditions imply (QC) and (4.17). In fact, they are equivalent as seen in [23]. This geometric interpretation is a key ingredient in showing regularity in the optimal transport problem in the vein of Caffarelli's classical work [7] (see [13, 16]).

5. Strong concavity of Kantorovich's functional

In this section we establish the strong concavity of Kantorovich's functional Φ over some suitable domain of $\mathbb{R}^Y$. As explained in the introduction, Φ is invariant under addition of a constant, so we must restrict ourselves to the orthogonal complement E_Y of the space of constant functions. Moreover, we will consider the set $\mathcal{K}^\varepsilon$ defined by (4.1), which can be thought of as the space of strictly c-concave functions. Recall that the conditions (Reg), (Twist), (QC), and (PW) are defined in the introduction, on pages 2604, 2605, 2608, and 2609 respectively.

Theorem 5.1. *Assume* (Reg), (Twist), *and* (QC). *Let X be a compact, c-convex subset of Ω, and ρ be a continuous probability density on X satisfying* (PW). *Then*

$$\forall \psi \in \mathcal{K}^\varepsilon,\ \forall v \in E_Y, \quad \langle \mathrm{D}^2\Phi(\psi)v \mid v\rangle \leq -C \cdot \varepsilon^3 \|v\|^2,$$

where C is a positive constant defined in (5.9), *and depends on $\|\rho\|_\infty$, $\mathcal{H}_g^{d-1}(\partial X)$, and $C_{\exp}$, C_∇, and $\varepsilon_{\mathrm{tw}}$ from Remark* 4.1.

Remark 5.1. Note that the upper bound on the largest non-zero eigenvalue of $\mathrm{D}^2\Phi(\psi)$ decreases as N grows to infinity, since ε is of the order of $1/N$. A possible place for improvement is the reverse isoperimetric inequality stated in (5.6). Currently, we are vastly overestimating the size of the boundary of a Laguerre cell by bounding it by the area of the boundary of the whole domain; additionally we are bounding the density ρ by its supremum, and paying in terms of the constant $C_{\exp}$. Note that (5.6) in its current form can *never* turn into equality, even for constant density ρ and the quadratic cost function where $C_{\exp} = 1$, as equality would only happen for a Laguerre cell that occupied the whole domain X, which cannot happen as *all* Laguerre cells have non-zero mass. To improve the inequality, one could try to control the anisotropy of Laguerre cells and bound the area of the boundary of a cell by some fraction of the area of ∂X; however, this would require assumptions on the distribution of the points Y and on $\nu \in \mathcal{P}(Y)$. We believe that such an upper bound on the anisotropy of Laguerre cells would be interesting in itself, and heuristically seems feasible as a discrete analogue of the regularity results for optimal transport (interpreting the Laguerre cells associated to $\psi : Y \to \mathbb{R}$ as the c^*-subdifferentials of ψ).

Remark 5.2. Note that unlike the domain X, the support of the density ρ need not be c-convex. In Appendix A we provide an example of a radial measure on $\mathbb{R}^d$ whose support is an annulus (hence is not simply connected) but whose Poincaré–Wirtinger constant C_{pw} is nonetheless positive.

The end of the section is devoted to the proof of Theorem 5.1. It relies on the fact that $-\mathrm{D}^2\Phi(\psi)$ can be regarded as the Laplacian matrix of a weighted graph on Y, whose first non-zero eigenvalue can be controlled from below using the Cheeger constant of the weighted graph. In turn, this weighted Cheeger constant can be controlled using the Poincaré–Wirtinger inequality.

5.1. Poincaré inequality and continuous Cheeger constant

We start by proving that the finiteness of the Poincaré–Wirtinger constant of the weighted domain (X, ρ) implies the positivity of the weighted Cheeger constant, defined in (5.1). Below, a *Lipschitz domain* is the closure of an open set with Lipschitz boundary.

Lemma 5.2. *Assume* (QC) *and that X is compact and c-convex. Then*

(i) *X is a Lipschitz domain,*
(ii) *for any $\psi \in \mathcal{K}^+$ and y in Y, $\mathrm{Lag}_y(\psi) \cap X$ is a Lipschitz domain.*

Proof. By assumption, for any $y \in Y$ one can write $X = \exp_y^c(X_y)$ where X_y is a bounded convex subset of $\mathbb{R}^d$ which must have non-empty interior since it supports an absolutely continuous probability measure. Moreover, the map $\exp_y^c$ is a diffeomorphism, hence is bi-Lipschitz. This implies (i), while (ii) follows from exactly the same arguments, where we have to remember that $\rho(\mathrm{Lag}_y(\psi)) > 0$. □

Given a Lipschitz domain A of X we write, slightly abusing notation,

$$|\partial A|_\rho := \int_{\partial A \cap \mathrm{int}(X)} \rho(x)\, \mathrm{d}\mathcal{H}_g^{d-1}(x) \quad \text{and} \quad |A|_\rho := \int_A \rho(x)\, \mathrm{d}\mathcal{H}_g^d(x).$$

Lemma 5.3. *Let X be a compact domain in Ω and let ρ in $\mathcal{C}^0(X)$ be a probability density with finite Poincaré–Wirtinger constant C_{pw}. Then the weighted Cheeger constant of (X, ρ) is positive, that is,*

$$\mathrm{h}(\rho) := \inf_{A \subseteq X} \frac{|\partial A|_\rho}{\min(|A|_\rho, |X \setminus A|_\rho)} \geq \frac{2}{C_{\mathrm{pw}}}, \tag{5.1}$$

where the infimum is taken over Lipschitz domains $A \subseteq \mathrm{int}(X)$ whose boundary has finite $\mathcal{H}_g^{d-1}$-measure.

The proof is based on properties of functions with bounded variation. For more details on this topic, we refer the reader to [4]. Although the discussion there is on Euclidean spaces, the relevant results easily extend to the Riemannian case, as $\exp_y^c$ serves as a global coordinate system on all of Ω.

Proof of Lemma 5.3. Let A be a Lipschitz domain in $\mathrm{int}(X)$. Since A has a Lipschitz boundary with finite area, its indicator function χ_A has bounded variation in $\mathrm{int}(X)$. By the density theorem [4, Theorem 10.1.2], there exists a sequence of $\mathcal{C}^1$ functions f_n on $\mathrm{int}(X)$ that converges to χ_A in the sense of intermediate convergence (whose definition is not important here). By (PW),

$$\|f_n - \mathbb{E}_\rho(f_n)\|_{L^1(\rho)} \leq C_{\mathrm{pw}} \|\nabla f_n\|_{L^1(\rho)}.$$

Since intermediate convergence is stronger than L^1 convergence, the continuity of ρ implies

$$\lim_{n\to\infty} \|f_n - \mathbb{E}_\rho(f_n)\|_{L^1(\rho)} = \|\chi_A - \mathbb{E}_\rho(\chi_A)\|_{L^1(\rho)} = 2|A|_\rho |X \setminus A|_\rho.$$

Note that we have used the fact that ρ is a probability measure, i.e. $\rho(X) = 1$. Proposition 10.1.2 of [4] implies that the total variation measure $|\mathrm{D} f_n|$ narrowly converges to $|\mathrm{D}\chi_A|$, which together with the continuity of ρ implies that $\int_\Omega |\mathrm{D} f_n| \rho \,\mathrm{d}\mathcal{H}^d_g$ converges to $\int_\Omega |\mathrm{D}\chi_A| \rho \,\mathrm{d}\mathcal{H}^d_g = |\partial A|_\rho$. The relation $|\mathrm{D} f_n| = \|\nabla f_n\|_g \mathrm{d}\mathcal{H}^d_g$ then gives

$$\lim_{n\to\infty} \|\nabla f_n\|_{L^1(\rho)} \leq |\partial A|_\rho.$$

Combining the previous equations leads to the desired inequality. □

5.2. *Cheeger constant of a graph*

The goal of this section is to give a lower bound of the second eigenvalue of $-\mathrm{D}^2\Phi(\psi)$ in terms of the Cheeger constant of the weighted graph induced by this matrix. An unoriented weighted graph can always be represented by its adjacency matrix $(w_{yz})_{(y,z)\in Y^2}$, a symmetric matrix with zero diagonal entries. We introduce a few definitions from graph theory, following the conventions of [14].

Definition 5.1. Let $(w_{yz})_{(y,z)\in Y^2}$ be a weighted graph over Y. The (weighted) *degree* of a vertex y is $d_y := \sum_{z\neq y} w_{yz}$. The (weighted) *Laplacian* is the matrix $(L_{yz})_{(y,z)\in Y^2}$ whose entries are $L_{yz} = -w_{yz}$ for $y \neq z$ and $L_{yy} = d_y$.

Definition 5.2. The *Cheeger constant* of a weighted graph $(w_{yz})_{(y,z)\in Y^2}$ over a point set Y is given by

$$\mathrm{h}(w) := \min_{S\subseteq Y} \frac{|\partial S|_w}{\min(|S|_w, |Y \setminus S|_w)}, \quad \text{where}$$
$$|\partial S|_w := \sum_{y\in S,\, z\notin S} w_{yz} \quad \text{and} \quad |S|_w := \sum_{y\in S} d_y.$$

The (weighted) Cheeger inequality bounds from below the first non-zero eigenvalue of the Laplacian of a weighted graph, denoted $\lambda(w)$, in terms of its Cheeger constant and its minimal degree. The formulation we use can be deduced from [14, Corollary 2.2] and from the inequality $1 - \sqrt{1 - x^2} \geq x^2/2$.

Theorem 5.4 (Cheeger inequality). $\lambda(w) \geq \frac{1}{2}\,\mathrm{h}^2(w) \cdot \min_{y\in Y} d_y$.

We now proceed to the proof of the main theorem of this section.

5.3. *Proof of Theorem 5.1*

Let ψ be a function in $\mathcal{K}^\varepsilon$ and consider the weighted graph $(w_{yz})_{(y,z)\in Y^2}$ given by

$$w_{yz} := -\frac{\partial^2 \Phi}{\partial \mathbb{1}_y \partial \mathbb{1}_z}(\psi) = \int_{\mathrm{Lag}_{y,z}(\psi)} \frac{\rho(x)}{\|\mathrm{D}_x c(x,y) - \mathrm{D}_x c(x,z)\|_g} \,\mathrm{d}\mathcal{H}_g^{d-1}(x)$$

for $y \neq z$ in Y, and with zero diagonal entries ($w_{yy} = 0$). In the formula above, we use the notation $\mathrm{Lag}_{y,z}(\psi) = \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$ for the facet between two Laguerre cells. By construction, the Laplacian matrix of this weighted graph is the Hessian matrix $-\mathrm{D}^2\Phi(\psi)$, so that Theorem 5.4 directly gives us a lower bound on the first non-zero eigenvalue of $-\mathrm{D}^2\Phi(\psi)$. To complete the proof, we need to bound the Cheeger constant and the minimum degree of the graph w from below.

Step 1. The goal here is to bound from below the discrete Cheeger constant $\mathrm{h}(w)$ in terms of the continuous weighted Cheeger constant $\mathrm{h}(\rho)$ and the constants introduced in (4.2). By definition of the constants $\varepsilon_{\mathrm{tw}}$ and C_∇, for any $y \neq z$ in Y one has

$$\varepsilon_{\mathrm{tw}} w_{yz} \leq |\mathrm{Lag}_{y,z}(\psi)|_\rho \leq 2C_\nabla w_{yz}. \tag{5.2}$$

Consider a subset S of Y, and let $A = \bigcup_{y\in S} \mathrm{Lag}_y(\psi)$. Then the intersection of the boundary of A with X is contained in a union of facets of Laguerre cells, namely

$$\partial A \cap \mathrm{int}(X) \subseteq \bigcup_{y\in S,\, z\notin S} \mathrm{Lag}_{y,z}(\psi). \tag{5.3}$$

The two inequalities (5.2) and (5.3) imply a lower bound on the numerator of the Cheeger constant:

$$|\partial A|_\rho \leq \sum_{y\in S,\, z\notin S} |\mathrm{Lag}_{y,z}(\psi)|_\rho \leq 2C_\nabla |\partial S|_w. \tag{5.4}$$

We now need to bound the denominator of the Cheeger constant from above, which requires controlling the weighted degrees d_y. Note that

$$d_y = \sum_{z\neq y} w_{yz} \leq \frac{1}{\varepsilon_{\mathrm{tw}}} \sum_{z\neq y} |\mathrm{Lag}_{y,z}(\psi)|_\rho \leq \frac{1}{\varepsilon_{\mathrm{tw}}} |\partial \,\mathrm{Lag}_y(\psi)|_\rho, \tag{5.5}$$

where the second inequality comes from the fact that the facets $\mathrm{Lag}_{y,z}(\psi)$ form a partition of the boundary $\partial \,\mathrm{Lag}_y(\psi) \cap \mathrm{int}(X)$ up to an $\mathcal{H}_g^{d-1}$-negligible set. To see that fact, it suffices to remark that in the exponential chart of y, the intersection of two distinct facets adjacent to y has a finite $\mathcal{H}_g^{d-2}$-measure, as implied by Proposition 3.2.

In order to apply the (continuous) Cheeger inequality, we need to replace the weighted area of the boundaries of Laguerre cells in (5.5) by the weighted volume of the cells. We have

$$\begin{aligned}
\mathcal{H}_g^{d-1}(\partial \,\mathrm{Lag}_y(\psi)) &\leq C_{\exp}^{d-1} \mathcal{H}^{d-1}((\exp_y^c)^{-1} \partial \,\mathrm{Lag}_y(\psi)) \\
&\leq C_{\exp}^{d-1} \mathcal{H}^{d-1}(\partial X_y) \leq C_{\exp}^{2(d-1)} \mathcal{H}_g^{d-1}(\partial X).
\end{aligned}$$

The first and third inequalities use the definition of the bi-Lipschitz constant C_{exp} of the exponential map, while the second inequality uses the monotonicity of the $\mathcal{H}^{d-1}$-measure of the boundary of a convex set with respect to inclusion (see [32, p. 211]). By the assumption $|\text{Lag}_y(\psi)|_\rho \geq \varepsilon$, this gives us a (rather crude) reverse isoperimetric inequality

$$\begin{aligned} |\partial \text{Lag}_y(\psi)|_\rho &\leq \|\rho\|_\infty \mathcal{H}_g^{d-1}(\partial \text{Lag}_y(\psi)) \\ &\leq \frac{\|\rho\|_\infty}{\varepsilon} C_{\text{exp}}^{2(d-1)} \mathcal{H}_g^{d-1}(\partial X) |\text{Lag}_y(\psi)|_\rho. \end{aligned} \tag{5.6}$$

We remark here that the above inequality is never sharp (see also Remark 5.1). Combining (5.5), (5.6) and $|A|_\rho = \sum_{y \in S} |\text{Lag}_y(\psi)|_\rho$ we obtain

$$|S|_w = \sum_{y \in S} d_y \leq \frac{1}{\varepsilon} \frac{\|\rho\|_\infty C_{\text{exp}}^{2(d-1)} \mathcal{H}_g^{d-1}(\partial X)}{\varepsilon_{\text{tw}}} |A|_\rho.$$

The same inequality holds for the complement $|X \setminus S|$. We combine the previous inequality with (5.4) and with Lemma 5.3 to get a lower bound on the Cheeger constant:

$$\mathrm{h}(w) \geq \frac{\varepsilon_{\text{tw}} \varepsilon}{C_{\text{exp}}^{2(d-1)} C_\nabla \mathcal{H}_g^{d-1}(\partial X) \|\rho\|_\infty C_{\text{pw}}}. \tag{5.7}$$

Note that in order to apply Lemma 5.3 we implicitly used the fact that A is a Lipschitz domain (as a finite union of Lipschitz domains, see Lemma 5.2) whose boundary has finite $\mathcal{H}_g^{d-1}$-measure (by (5.6)).

Step 2. In order to apply the Cheeger inequality, we still need to bound from below the weighted degree d_y. By (5.2) one has, in view of the crucial fact that $|\partial \text{Lag}_y(\psi)|_\rho$ is the measure of $\partial \text{Lag}_y(\psi) \cap \text{int}(X)$,

$$d_y = \sum_{z \neq y} w_{yz} \geq \frac{1}{2C_\nabla} \sum_{z \neq y} |\text{Lag}_{y,z}(\psi)|_\rho \geq \frac{1}{2C_\nabla} |\partial \text{Lag}_y(\psi)|_\rho.$$

Taking $A = \text{Lag}_y(\psi)$ in the definition of the weighted Cheeger constant $\mathrm{h}(\rho)$ in Lemma 5.3, one gets

$$|\partial \text{Lag}_y(\psi)|_\rho \geq \mathrm{h}(\rho) \min\big(|\text{Lag}_y(\psi)|_\rho, |X \setminus \text{Lag}_y(\psi)|_\rho\big) \geq \mathrm{h}(\rho)\varepsilon.$$

The last inequality comes from the assumption that each Laguerre cell has a mass greater than ε and that $X \setminus \text{Lag}_y(\psi)$ also contains a Laguerre cell (except for the trivial case where Y is a singleton). We deduce

$$d_y \geq \frac{\varepsilon}{C_\nabla C_{\text{pw}}}. \tag{5.8}$$

Step 3. Combining the Cheeger inequality with (5.7) and (5.8) we have $\lambda(w) \geq C\varepsilon^3$ where

$$C := \frac{\varepsilon_{\mathrm{tw}}^2}{2C_{\exp}^{4(d-1)} C_{\nabla}^3 (\mathcal{H}_g^{d-1}(\partial X))^2 \|\rho\|_\infty^2 C_{\mathrm{pw}}^3}. \tag{5.9}$$

Since the graph induced by the Hessian is connected, the kernel of $-\mathrm{D}^2\Phi(\psi)$ is equal to the space of constant functions over Y, implying that $\mathrm{Ker}(-\mathrm{D}^2\Phi(\psi))^\perp = E_Y$. Then, using the variational characterization of the first non-zero eigenvalue of the Laplacian matrix, we get

$$C\varepsilon^3 \leq \lambda(w) = \min_{v \in E_Y} \frac{\langle -\mathrm{D}^2\Phi(\psi) \mid v\rangle}{\|v\|^2}. \qquad \square$$

6. Convergence of the damped Newton algorithm

The goal of this section is to show the convergence of the damped Newton algorithm for semi-discrete optimal transport. This follows in fact from a more general result. We establish in Section 6.1 the convergence of the damped Newton algorithm (Algorithm 1) under general assumptions on the functional. We finally apply this algorithm to the semi-discrete optimal transport problem, using the intermediate results (regularity and strict concavity of the Kantorovich functional) proven in Sections 4 and 5.

6.1. General damped Newton algorithm

Recall that Y is a finite set and we denote by $\mathbb{R}^Y$ the space of real functions on Y. We consider $\mathcal{P}(Y)$, the space of probability measures on Y, as a subset of $\mathbb{R}^Y$. Finally, we denote by E_Y the space of functions on Y that sum to zero. In this section, we show that Algorithm 1 can be used to solve non-linear equations $G(\psi) = \mu$ where $\mu \in \mathcal{P}(Y)$ and the map $G : \mathbb{R}^Y \to \mathcal{P}(Y)$ satisfies some regularity and monotonicity assumptions.

Proposition 6.1. *Let G be a functional from $\mathbb{R}^Y$ to $\mathcal{P}(Y)$ which is invariant under addition of a constant. Let $G(\psi) = \sum_{y\in Y} G_y(\psi)\mathbb{1}_y$ and*

$$\mathcal{K}^\varepsilon = \{\psi \in \mathbb{R}^Y \mid \forall y \in Y,\ G_y(\psi) \geq \varepsilon\},$$

and assume that G has the following properties:

(i) (Regularity) *For every positive ε, G is $\mathcal{C}^{1,\alpha}$ on $\mathcal{K}^\varepsilon$. Let L_ε be the smallest constant such that*

$$\forall \varphi \neq \psi \in \mathcal{K}^\varepsilon, \qquad \frac{\|G(\varphi) - G(\psi)\|}{\|\varphi - \psi\|} + \frac{\|\mathrm{D}G(\varphi) - \mathrm{D}G(\psi)\|}{\|\varphi - \psi\|^\alpha} \leq L_\varepsilon.$$

(ii) (Uniform monotonicity) *For every $\varepsilon > 0$, there exists a positive constant κ_ε such that G is κ_ε-uniformly monotone on $\mathcal{K}^\varepsilon \cap E_Y$:*

$$\forall \psi \in \mathcal{K}^\varepsilon, \forall v \in E_Y, \qquad \langle v \mid \mathrm{D}G(\psi)v\rangle \geq \kappa_\varepsilon \|v\|^2.$$

Now, let $\mu \in \mathcal{P}(Y)$ and let ψ_0 be a function on Y such that the constant ε_0 defined in (1.6) is positive. Set $\kappa := \min(\kappa_{\varepsilon_0/2}, 1)$ and $L := \max(L_{\varepsilon_0/2}, 1)$. Then the iterates (ψ_k) of Algorithm 1 satisfy

$$\|G(\psi_{k+1}) - \mu\| \leq (1 - \overline{\tau}_k/2)\|G(\psi_k) - \mu\|, \quad \text{where}$$
$$\overline{\tau}_k := \min\left(\frac{\kappa^{1+1/\alpha}\varepsilon}{d^{1/\alpha}L^{1/\alpha}\|G(\psi_k) - \mu\|}, 1\right). \tag{6.1}$$

In addition, as soon as $\overline{\tau}_k = 1$ one has

$$\|G(\psi_{k+1}) - \mu\| \leq \frac{L\|G(\psi_k) - \mu\|^{1+\alpha}}{\kappa^{1+\alpha}}.$$

In particular, the damped Newton algorithm converges globally with linear speed and locally with superlinear speed (quadratic speed if $\alpha = 1$).

Proof. We set $\varepsilon := \varepsilon_0$, $L := \max(L_{\varepsilon/2}, 1)$ and $\kappa := \min(\kappa_{\varepsilon/2}, 1)$. First, we remark that for every $\psi \in \mathcal{K}^{\varepsilon/2}$, the pseudo-inverse $\mathrm{D}G^+(\psi)$ maps the subspace E_Y to itself. The uniform monotonicity of G therefore implies that $\|\mathrm{D}G^+(\psi)\| \leq 1/\kappa$, where $\|\cdot\|$ is the operator norm on $\mathbb{R}^Y$.

We start by the analysis of a single iteration of the algorithm. We let $\psi := \psi_k \in \mathcal{K}^\varepsilon$, define $v := \mathrm{D}G(\psi)^+(G(\psi) - \mu)$ and $\psi_\tau := \psi - \tau v$. Since the pseudo-inverse $\mathrm{D}G^+(\psi)$ is $1/\kappa$-Lipschitz, one has $\|v\| \leq \|G(\psi) - \mu\|/\kappa$. Now let τ_1 be the largest time before the curve ψ_σ leaves $\mathcal{K}^{\varepsilon/2}$. In particular, ψ_{τ_1} lies at the boundary of $\mathcal{K}^{\varepsilon/2}$, meaning that there must exist a point y in Y such that $G_y(\psi_{\tau_1}) = \varepsilon/2$. This implies that $\|G(\psi_{\tau_1}) - G(\psi)\| \geq \varepsilon/2$, and using the Lipschitz bound on G we obtain a lower bound on τ_1:

$$\frac{\varepsilon}{2} \leq \|G(\psi_{\tau_1}) - G(\psi)\| \leq L\tau_1\|v\| \leq \frac{L\tau_1}{\kappa}\|G(\psi) - \mu\|.$$

This implies that τ_1 is necessarily larger than $\kappa\varepsilon/(2L\|G(\psi) - \mu\|)$. We have now established that the curve $\tau \mapsto \psi_\tau$ remains in $\mathcal{K}^{\varepsilon/2}$ before time τ_1, implying that the function $[0, \tau_1] \ni \tau \mapsto G(\psi_\tau)$ is uniformly $\mathcal{C}^{1,\alpha}$. Applying Taylor's formula we get

$$G(\psi_\tau) = G\big(\psi - \tau \mathrm{D}G(\psi)^+(G(\psi) - \mu)\big) = (1 - \tau)G(\psi) + \tau\mu + R(\tau), \tag{6.2}$$

where, on account of $v = \mathrm{D}G(\psi)^+(G(\psi) - \mu)$ and the α-Hölder property for $\mathrm{D}G$,

$$\begin{aligned}\|R(\tau)\| &= \left\|\int_0^\tau (\mathrm{D}G(\psi_\sigma) - \mathrm{D}G(\psi))v \,\mathrm{d}\sigma\right\| \\ &\leq \frac{L}{\alpha+1}\tau^{\alpha+1}\|v\|^{\alpha+1} \leq \frac{L\|G(\psi) - \mu\|^{1+\alpha}}{\kappa^{1+\alpha}}\tau^{1+\alpha}.\end{aligned} \tag{6.3}$$

For every $y \in Y$, using $\mu_y \geq 2\varepsilon$ (by (1.6)) and $G_y(\psi) \geq \varepsilon$, one gets

$$G_y(\psi_\tau) \geq (1 - \tau)G_y(\psi) + \tau\mu_y + R_y(\tau) \geq (1 + \tau)\varepsilon - \|R(\tau)\|.$$

If τ is chosen such that $\|R(\tau)\| \leq \tau\varepsilon$ we will have $G_y(\psi_\tau) \geq \varepsilon$ for all y in Y and therefore ψ_τ will belong to $\mathcal{K}^\varepsilon$. Thanks to our estimate on $R(\tau)$ this will be true provided that

$$\tau \leq \tau_2 := \min\left(\tau_1, \frac{\kappa^{1+1/\alpha}\varepsilon^{1/\alpha}}{L^{1/\alpha}\|G(\psi)-\mu\|^{1+1/\alpha}}\right).$$

Finally, we establish the second inequality required by Step 2 of the algorithm. To do that, we subtract μ from both sides in (6.2) to obtain

$$G(\psi_\tau) - \mu = (1-\tau)(G(\psi)-\mu) + R(\tau). \tag{6.4}$$

In order to get $\|G(\psi_\tau)-\mu\| \leq (1-\tau/2)\|G(\psi)-\mu\|$, it is sufficient to establish that $\|R(\tau)\| \leq \tau/2\|G(\psi)-\mu\|$. Using the estimation on $\|R(\tau)\|$ again, we see that it suffices to take

$$\tau \leq \tau_3 := \min\left(\tau_2, \frac{\kappa^{1+1/\alpha}}{L^{1/\alpha}\|G(\psi)-\mu\|2^{1/\alpha}}, 1\right).$$

Finally, using $L \geq 1$, $\kappa \leq 1$ and $\|G(\psi)-\mu\| \leq d$ (since $G(\psi)$ and μ are probability measures), we can establish that $\tau_3 \geq \overline{\tau}_k$ where $\overline{\tau}_k$ is defined in (6.1). This ensures the first estimate on the improvement of the error between two successive steps.

By this estimate, there exists k_0 such that $\overline{\tau}_k = 1$ for $k \geq k_0$. Then one can use (6.4) to get $\|G(\psi_{k+1})-\mu\| \leq \|R(\tau)\|$. We obtain the second estimate of the theorem by plugging in (6.3). □

6.2. *Proof of Theorem 1.5*

Proposition 6.1 can be directly applied to the gradient of the Kantorovich functional, or more precisely to

$$G(\psi) := \sum_{y\in Y} \rho(\mathrm{Lag}_y(\psi))\mathbb{1}_y = \nabla\Phi(\psi) + \mu.$$

In that case, the set $\mathcal{K}^\varepsilon$ is given by

$$\mathcal{K}^\varepsilon = \{\psi \in \mathbb{R}^Y \mid \forall y \in Y,\ \rho(\mathrm{Lag}_y(\psi)) \geq \varepsilon\}.$$

We have assumed that the probability density ρ is in $\mathcal{C}^{0,\alpha}(X)$ where X is a c-convex, compact subset of Ω. Then, by Theorem 4.1, for any $\varepsilon > 0$, the map G is uniformly $\mathcal{C}^{1,\alpha}$ over $\mathcal{K}^\varepsilon$. This ensures that the (Regularity) condition of Proposition 6.1 is satisfied. Furthermore, since we have also assumed that ρ satisfies a weighted Poincaré–Wirtinger inequality, we can apply Theorem 5.1 to see that the (Uniform monotonicity) hypothesis of Proposition 6.1 is also satisfied. Applying Proposition 6.1, we deduce the desired convergence rates for Algorithm 1.

6.3. *Numerical results*

We conclude the article with a numerical illustration of this algorithm, for the cost $c(x, y) = \|x - y\|^2$ and for a piecewise-linear density. The source density is piecewise-linear over a triangulation of [0, 3] with 18 triangles (displayed in Figure 3). It takes value 1 on the boundary $\partial[0, 3]^2$ and vanishes on the square $[1, 2]^2$. In particular, the support of this density is not simply connected and not convex. The target measure is uniform over a

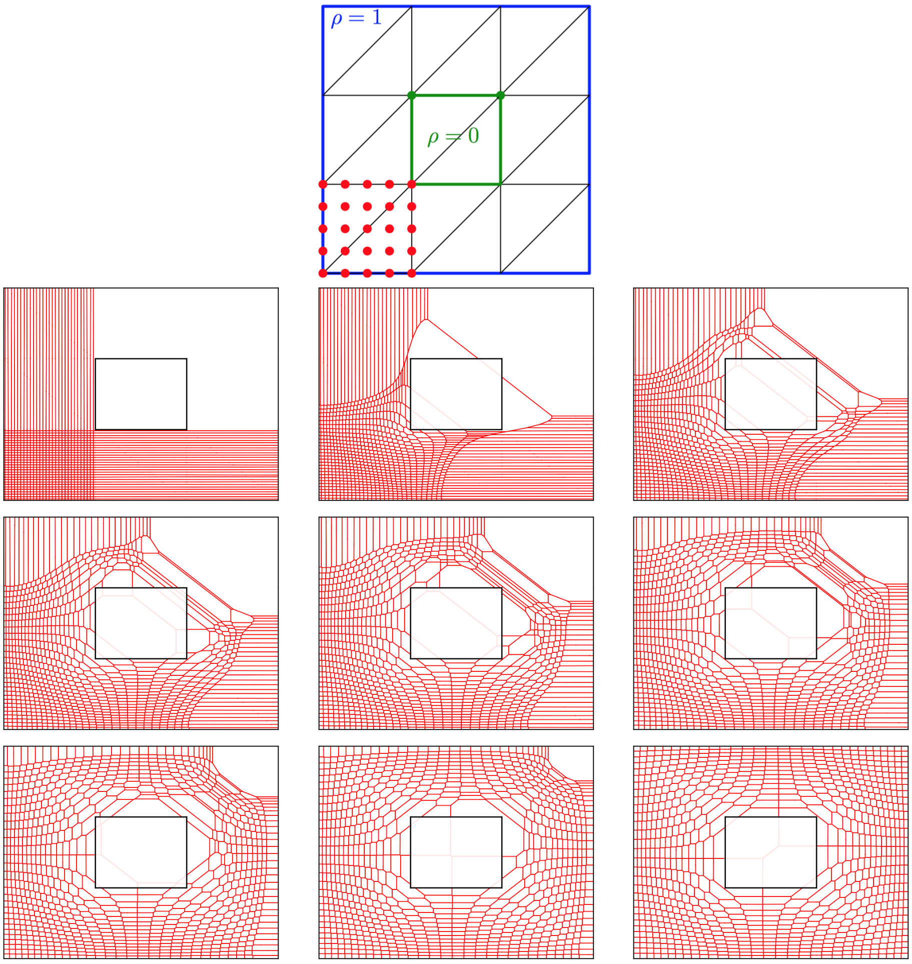

Fig. 3. Evolution of Laguerre cells during the execution of the damped Newton algorithm for semi-discrete optimal transport. Top: The source density ρ is piecewise linear over the domain $X = [0, 3]^3$ over the displayed triangulation: it takes value 1 on the boundary of the square $[0, 3]^2$ and 0 on the boundary of $[1, 2]^2$. The target measure is uniform over a 30^2 uniform grid in $[0, 1]^2$. Bottom: Laguerre cells at steps $k = 0, 2, 6, 9, 12, 15, 18, 21$ and 25.

uniform grid $\frac{1}{n-1}\{0, \dots, n-1\}^2$. Figure 3 displays the iterates of the Newton algorithm, which in this case takes 25 iterations to solve the optimal transport problem with an error equal to the numerical precision of the machine. The source code of this algorithm is publicly available.[1]

We finally note that recent progress in computational geometry would allow one to implement Algorithm 1 for the quadratic cost on $\mathbb{R}^3$, refining [22] or [12]. It should also be possible to deal with optimal transport problems arising from geometric optics, such as the far-field reflector problem [10], whose associated cost satisfies the Ma–Trudinger–Wang condition [24].

Appendix A. A weighted Poincaré–Wirtinger inequality

In this section, we provide an (almost) explicit example of a probability density on $\mathbb{R}^d$ whose support is an annulus, therefore not simply connected, but which still satisfies a weighted Poincaré–Wirtinger inequality.

Proposition A.1. *Let $0 < r < R$ and assume that $\overline{\rho} \in \mathcal{C}^0([0, R])$ is a probability density with $\overline{\rho} = 0$ on $[0, r]$ and $\overline{\rho}$ concave on $[r, R]$. Consider*

$$\rho(x) = \frac{1}{\|x\|^{d-1}\omega_{d-1}}\overline{\rho}(\|x\|) \quad \text{over } X := \mathrm{B}(0, R) \subseteq \mathbb{R}^d,$$

where ω_{d-1} is the volume of the unit sphere $\mathbb{S}^{d-1}$. Then ρ satisfies the weighted Poincaré–Wirtinger inequality (PW) for some positive constant.

The proof relies on two L^1-Poincaré–Wirtinger inequalities. The first inequality is the usual Poincaré–Wirtinger inequality on the sphere: given a $\mathcal{C}^1$ function f on $\mathbb{S}^{d-1}$, and $F_{d-1} := (1/\omega_{d-1}) \int_{\mathbb{S}^{d-1}} f(z)\,\mathrm{d}z$,

$$\int_{\mathbb{S}^{d-1}} |f(z) - F_{d-1}|\,\mathrm{d}\mathcal{H}^{d-1}(z) \le c_d \int_{\mathbb{S}^{d-1}} \|\nabla f(z)\|\,\mathrm{d}\mathcal{H}^{d-1}(z) \tag{A.1}$$

for some positive constant c_d. The second inequality is a Poincaré–Wirtinger inequality on the interval $[0, R]$ weighted by $\overline{\rho}$. Given a function $\overline{f}$ in $\mathcal{C}^1([0, R])$, and letting $F_1 := \int_0^R \overline{f}(r)\overline{\rho}(r)\,\mathrm{d}r / \int_0^R \overline{\rho}(r)\,\mathrm{d}r$, we have

$$\int_0^R |\overline{f}(r) - F_1|\overline{\rho}(r)\,\mathrm{d}r \le c_{\overline{\rho}} \int_0^R |\overline{f}'(r)|\overline{\rho}(r)\,\mathrm{d}r \tag{A.2}$$

for some positive constant $c_{\overline{\rho}}$ depending only on $\overline{\rho}$, as can be deduced from [3, Theorem 2.1] and from the concavity of $\overline{\rho}$ on $[r, R]$.

[1] https://github.com/mrgt/PyMongeAmpere.

Proof of Proposition A.1. We now proceed to the proof of the Poincaré–Wirtinger inequality for (X, ρ). Let $f : \mathrm{B}(0, R) \to \mathbb{R}$ be a function of class $\mathcal{C}^1$. By using polar coordinates and the definition of ρ, one has

$$F := \int_{\mathrm{B}(0,R)} f(x)\rho(x)\,\mathrm{d}\mathcal{H}^d(x) = \int_0^R \frac{1}{\omega_{d-1}r^{d-1}} \int_{\mathbb{S}^{d-1}(r)} f(z)\overline{\rho}(r)\,\mathrm{d}\mathcal{H}^{d-1}(z)\,\mathrm{d}r = \int_0^R \overline{f}(r)\overline{\rho}(r)\,\mathrm{d}r,$$

where $\overline{f}(r)$ is the mean value of f over the sphere $\mathbb{S}^{d-1}(r)$,

$$\overline{f}(r) = \frac{1}{\omega_{d-1}r^{d-1}} \int_{\mathbb{S}^{d-1}(r)} f(z)\,\mathrm{d}\mathcal{H}^{d-1}(z) = \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} f(rz)\,\mathrm{d}\mathcal{H}^{d-1}(z).$$

Using the triangle inequality and the relation between $\overline{\rho}$ and ρ we get

$$\begin{aligned}\int_{\mathrm{B}(0,R)} |f(x) - F|\rho(x)\,\mathrm{d}\mathcal{H}^d(x) &= \int_0^R \int_{\mathbb{S}^{d-1}(r)} |f(z) - F|\rho(z)\,\mathrm{d}\mathcal{H}^{d-1}(z)\,\mathrm{d}r \\ &\leq \int_0^R \overline{\rho}(r)|\overline{f}(r) - F|\,\mathrm{d}r + \int_0^R \frac{\overline{\rho}(r)}{r^{d-1}\omega_{d-1}} \int_{\mathbb{S}^{d-1}(r)} |f(z) - \overline{f}(r)|\,\mathrm{d}\mathcal{H}^{d-1}(z)\,\mathrm{d}r. \end{aligned} \tag{A.3}$$

We first deal with the second term on the right-hand side. Using the Poincaré–Wirtinger inequality (A.1) on the sphere, we have

$$\int_{\mathbb{S}^{d-1}(r)} |f(z) - \overline{f}(r)|\,\mathrm{d}\mathcal{H}^{d-1}(z) \leq c_d \int_{\mathbb{S}^{d-1}(r)} \|\nabla f(z)|_{z^\perp}\|\,\mathrm{d}\mathcal{H}^{d-1}(z),$$

where $\nabla f(z)|_{z^\perp}$ is the orthogonal projection of the gradient on the tangent plane $\{z\}^\perp$, so that

$$\int_0^R \frac{\overline{\rho}(r)}{r^{d-1}\omega_{d-1}} \int_{\mathbb{S}^{d-1}(r)} |f(z) - \overline{f}(r)|\,\mathrm{d}\mathcal{H}^{d-1}(z) \leq c_d \int_{\mathrm{B}(0,r)} \|\nabla f(x)|_{x^\perp}\|\rho(x)\,\mathrm{d}\mathcal{H}^d(x). \tag{A.4}$$

By the calculation of F above, we see that F is also the mean value of $\overline{f}$ weighted by $\overline{\rho}$. We can therefore control the first term of the upper bound of (A.3) using the Poincaré–Wirtinger inequality (A.2) on the interval:

$$\int_0^R \overline{\rho}(r)|\overline{f}(r) - F|\,\mathrm{d}r \leq c_{\overline{\rho}} \int_0^R |\overline{f}'(r)|\overline{\rho}(r)\,\mathrm{d}r.$$

Now, notice that

$$\overline{f}'(r) = \lim_{h\to 0} \frac{\overline{f}(r+h) - \overline{f}(r)}{h} = \lim_{h\to 0} \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \frac{f((r+h)z) - f(rz)}{h}\,\mathrm{d}\mathcal{H}^{d-1}(z),$$

from which we deduce

$$|\overline{f}'(r)| \leq \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \left| \frac{\partial f}{\partial r}(rz) \right| \mathrm{d}\mathcal{H}^{d-1}(z)$$
$$= \frac{1}{\omega_{d-1} r^{d-1}} \int_{\mathbb{S}^{d-1}(r)} \left| \left\langle \nabla f(z) \,\middle|\, \frac{z}{r} \right\rangle \right| \mathrm{d}\mathcal{H}^{d-1}(z).$$

Integrating this inequality shows that

$$\int_0^R \overline{\rho}(r) |\overline{f}(r) - F| \,\mathrm{d}r \leq c_{\overline{\rho}} \int_0^R \frac{\overline{\rho}(r)}{\omega_{d-1} r^{d-1}} \int_{\mathbb{S}^{d-1}(r)} \left| \left\langle \nabla f(z) \,\middle|\, \frac{z}{r} \right\rangle \right| \mathrm{d}\mathcal{H}^{d-1}(z)$$
$$= c_{\overline{\rho}} \int_{\mathrm{B}(0,R)} \left| \left\langle \nabla f(x) \,\middle|\, \frac{x}{\|x\|} \right\rangle \right| \rho(x) \,\mathrm{d}\mathcal{H}^{d}(x). \tag{A.5}$$

From the simple inequality $(a+b)^2 \leq 2(a^2+b^2)$, we get

$$\left| \left\langle \nabla f(x) \,\middle|\, \frac{x}{\|x\|} \right\rangle \right| + \|\nabla f(x)|_{x^\perp}\| \leq \sqrt{2}\, \|\nabla f(x)\|.$$

Using the bounds (A.4) and (A.5) in (A.3), we get the desired inequality:

$$\int_{\mathrm{B}(0,R)} |f(x) - F| \rho(x) \,\mathrm{d}\mathcal{H}^d(x) \leq \sqrt{2}(c_d + c_{\overline{\rho}}) \int_{\mathrm{B}(0,R)} \|\nabla f(x)\| \rho(x) \,\mathrm{d}\mathcal{H}^d(x). \qquad \square$$

Appendix B. Proof of Theorem 3.1

B.1. Existence of partial derivatives

Without loss of generality, we assume that $\boldsymbol{\lambda}_0 = \mathbf{0}$. We start the proof of Theorem 3.1 by showing the existence of partial derivatives of the map $\hat{G}$. In this section, we denote by $e_1, \ldots, e_N$ the canonical basis of $\mathbb{R}^N$. We start by rewriting the finite difference defining the partial derivative of $\hat{G}$ in direction e_i using the coarea formula. Fix $\|\boldsymbol{\lambda}\| < T_{\mathrm{tr}}$. For $t > 0$, one has

$$\frac{1}{t}(\hat{G}(\boldsymbol{\lambda} + t\boldsymbol{e}_i) - \hat{G}(\boldsymbol{\lambda})) = \frac{1}{t} \int_{K(\boldsymbol{\lambda}+t\boldsymbol{e}_i)\setminus K(\boldsymbol{\lambda})} \hat{\rho}(x) \,\mathrm{d}\mathcal{H}^d(x) = \frac{1}{t} \int_{\lambda_i}^{\lambda_i + t} \hat{g}(s) \,\mathrm{d}s, \tag{B.1}$$

where the function $\hat{g}$ is defined by

$$\hat{g}(s) := \int_{\bigcap_{j\neq i} K_j(\lambda_j) \cap f_i^{-1}(s)} \frac{\hat{\rho}(x)}{\|\nabla f_i(x)\|} \,\mathrm{d}\mathcal{H}^{d-1}(x). \tag{B.2}$$

The same reasoning also holds for $t < 0$. We now claim that $\hat{g}$ is continuous on some interval around λ_i, which by (B.1) and the Fundamental Theorem of Calculus will imply that the limit as $t \to 0$ of (B.1) exists and is equal to $\hat{g}(\boldsymbol{\lambda})$, thus establishing the formula (3.3). The continuity of $\hat{g}$ follows from the next proposition, which is formulated in a slightly more general way.

Proposition B.1. *Let σ be a continuous non-negative function on $\hat{X}$ and let ω be the modulus of continuity of σ. Given any vector $\boldsymbol{\lambda}$ in $\mathbb{R}^N$ with $\|\boldsymbol{\lambda}\|_\infty \leq T_{\mathrm{tr}}$, consider the function*

$$h : \mathbb{R} \ni s \mapsto \int_{L\cap S_s} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x),$$

where $L := \bigcap_{j\neq i} K_j(\lambda_j)$ and $S_s := f_i^{-1}(s)$. Then h is uniformly continuous on $[-T_{\mathrm{tr}}, T_{\mathrm{tr}}]$ and has modulus of continuity

$$\omega_h(\delta) = C_1 \cdot (\omega(C_2\delta) + |\delta|), \tag{B.3}$$

where the constants only depend on $\|f_i\|_{\mathcal{C}^{1,1}}$, $\mathrm{diam}(\hat{X})$, $\varepsilon_{\mathrm{nd}}$, $\varepsilon_{\mathrm{tr}}$ and $\|\sigma\|_\infty$.

Taking $\sigma = \hat{\rho}/\|\nabla f_i\|$ in the previous proposition, which is continuous using the non-degeneracy condition (ND) and the assumption $f_i \in \mathcal{C}^{1,1}(\hat{X})$, we see that the function $\hat{g}$ defined by (B.2) is continuous. This implies the existence of partial derivatives and establishes formula (3.3). The proof of Proposition B.1 requires the following lemma.

Lemma B.2. *Assume that the functions $f_i : \hat{X} \to \mathbb{R}$ satisfy (ND). Then, for every $i \in \{1, \ldots, N\}$, there exists a map $\Phi_i : \hat{X} \times \mathbb{R} \to \mathbb{R}^d$ such that:*

(i) *For any (x, t) in $\hat{X} \times \mathbb{R}$ such that the curve $\Phi_i(x, [0, t])$ remains in $\hat{X}$, one has $f_i(\Phi_i(x, t)) = f_i(x) + t$.*
(ii) *For all $x, y \in \hat{X}$ and $t \in \mathbb{R}$,*

$$\|\Phi_i(x, t) - \Phi_i(x, s)\| \leq |t - s|/\varepsilon_{\mathrm{nd}}, \tag{B.4}$$
$$\|\Phi_i(x, t) - \Phi_i(y, t)\| \leq \exp(C_\Phi|t|)\|x - y\|, \tag{B.5}$$

where $C_\Phi := 3C_L/\varepsilon_{\mathrm{nd}}^2$.

Proof. We consider the vector field $V_i^0(x) = \nabla f_i(x)/\|\nabla f_i(x)\|^2$ on $\hat{X}$, which satisfies $\|V_i^0\|_\infty \leq 1/\varepsilon_{\mathrm{nd}}$ and whose Lipschitz constant is bounded by C_Φ. This vector field is extended to $\mathbb{R}^d$ using the orthogonal projection on $\hat{X}$, denoted $p_{\hat{X}}$:

$$\forall x \in \mathbb{R}^d, \qquad V_i(x) := V_i^0(p_{\hat{X}}(x)).$$

By convexity of $\hat{X}$, the map $p_{\hat{X}}$ is 1-Lipschitz. This implies that the Lipschitz constant of V_i is also bounded by C_Φ. We let Φ_i be the flow induced by this vector field, which exists for all time since V_i is bounded and uniformly Lipschitz on all of $\mathbb{R}^d$. The inequality (B.4) follows from the definition of integral curves and the bound on $\|V_i\|$. Any integral curve $\gamma : [0, T] \to \mathbb{R}^d$ of V_i which remains in $\hat{X}$ satisfies

$$\begin{aligned} f_i(\gamma(t)) &= f_i(\gamma(0)) + \int_0^t \langle \gamma'(s) \mid \nabla f_i(\gamma(s))\rangle\, \mathrm{d}s \\ &= f_i(\gamma(0)) + \int_0^t \langle V_i(\gamma(s)) \mid \nabla f_i(\gamma(s))\rangle\, \mathrm{d}s = f_i(\gamma(0)) + t, \end{aligned}$$

thus establishing (i). The inequality (B.5) follows from the bound on the Lipschitz constant of V_i and from Gronwall's lemma. □

Proof of Proposition B.1. Let t, s be small enough so that the transversality condition (T) holds (that is, $t, s \in [-T_{\mathrm{tr}}, T_{\mathrm{tr}}]$). We assume that $t < s$ in order to fix the signs of some expressions. We consider the following partition of the facet $S_t \cap L$, whose geometric meaning is illustrated in Figure 4:

$$A_t = \{x \in S_t \cap L \mid \Phi_i(x, [0, s-t]) \subseteq L\},$$
$$B_t = \{x \in S_t \cap L \mid \exists u \in [0, s-t),\ \Phi_i(x, u) \in \partial L\}.$$

Similarly, we define

$$A_s = \{x \in S_s \cap L \mid \Phi_i(x, [t-s, 0]) \subseteq L\},$$
$$B_s = \{x \in S_s \cap L \mid \exists u \in (t-s, 0],\ \Phi_i(x, u) \in \partial L\}.$$

Recall that by definition,

$$h(t) = \int_{A_t} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x) + \int_{B_t} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x), \tag{B.6}$$

where the integral is with respect to the $(d-1)$-dimensional Hausdorff measure. Our strategy to show the continuity of h is to prove that the first terms in the sums defining $h(t)$ and $h(s)$ in (B.6) are close, namely

$$\left|\int_{A_t} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x) - \int_{A_s} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x)\right| \leq C_3 \cdot \big(|s-t| + \omega(C|s-t|)\big), \tag{B.7}$$

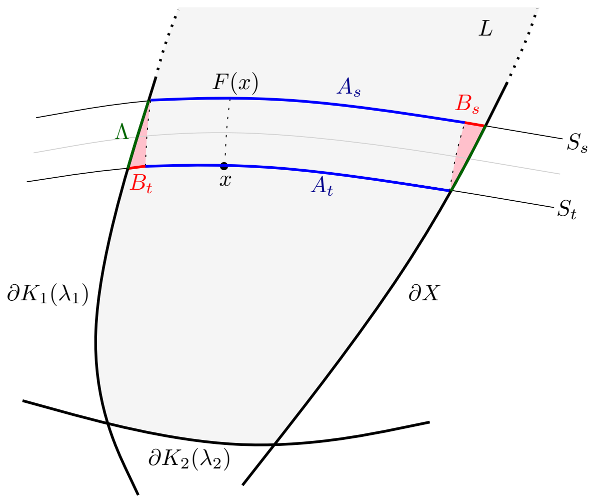

Fig. 4. Illustration of the proof of Proposition B.1.

and then that the terms involving B_t, B_s are small (recall that both sets depend on t *and* s):

$$\int_{B_t} |\sigma(x)| \, \mathrm{d}\mathcal{H}^{d-1}(x) + \int_{B_s} |\sigma(x)| \, \mathrm{d}\mathcal{H}^{d-1}(x) \le C_4 \cdot |s-t|. \tag{B.8}$$

The combination of the estimates (B.7) and (B.8) implies the desired inequality (B.3). We now turn to the proof of these estimates, and that the constants C_3 and C_4 in these estimates depend on $\|f_i\|_{\mathcal{C}^{1,1}}$, $\mathrm{diam}(\hat{X})$, $\varepsilon_{\mathrm{nd}}$, $\varepsilon_{\mathrm{tr}}$ and $\|\sigma\|_\infty$.

Proof of (B.7). By Lemma B.2(i), for any point x in A_t one has $f_i(\Phi_i(x, s-t)) = s$, so that the map $F(x) := \Phi_i(x, s-t)$ induces a bijection between the sets A_t and A_s. As a consequence of (B.5), the restriction of F to A_t is a bi-Lipschitz bijection between the sets A_t and A_s, with Lipschitz constant

$$\max\{\|F^{-1}\|_{\mathrm{Lip}(A_s)}, \|F\|_{\mathrm{Lip}(A_t)}\} \le \exp(C_\Phi |s-t|).$$

Using a Lipschitz change of variable formula, we get

$$\begin{aligned} \int_{A_t} \sigma(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) &= \int_{F^{-1}(A_s)} \sigma(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) \\ &\le \|F^{-1}\|_{\mathrm{Lip}(A_s)}^{d-1} \int_{A_s} \sigma(F^{-1}(x)) \, \mathrm{d}\mathcal{H}^{d-1}(x) \\ &\le \exp\big(C_\Phi (d-1)|s-t|\big) \int_{A_s} \sigma(F^{-1}(x)) \, \mathrm{d}\mathcal{H}^{d-1}(x). \end{aligned} \tag{B.9}$$

By definition of the modulus of continuity and thanks to (B.4),

$$\begin{aligned} |\sigma(F^{-1}(x)) - \sigma(x)| &\le \omega(\|F^{-1}(x) - x\|) \\ &= \omega(\|\Phi(x, s-t) - x\|) \le \omega(|s-t|/\varepsilon_{\mathrm{nd}}). \end{aligned}$$

Integrating this inequality, we get

$$\begin{aligned} \int_{A_s} \sigma(F^{-1}(x)) \, \mathrm{d}\mathcal{H}^{d-1}(x) &\le \int_{A_s} \sigma(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) + \mathcal{H}^{d-1}(A_s) \omega(|s-t|/\varepsilon_{\mathrm{nd}}) \\ &\le \int_{A_s} \sigma(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) + \mathcal{H}^{d-1}(\hat{X}) \omega(|s-t|/\varepsilon_{\mathrm{nd}}), \end{aligned} \tag{B.10}$$

where the second inequality uses the monotonicity of the $(d-1)$-dimensional Hausdorff measure of the boundary of a convex set with respect to inclusion (see [32, p. 211]). Combining (B.9) and (B.10) we get

$$\begin{aligned} &\int_{A_t} \sigma(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) \\ &\quad \le \exp\big(C_\Phi (d-1)|s-t|\big) \left(\int_{A_s} \sigma(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) + \mathcal{H}^{d-1}(\hat{X}) \omega(|s-t|/\varepsilon_{\mathrm{nd}}) \right), \end{aligned}$$

so that

$$\begin{aligned}\int_{A_t}\sigma(x)\,\mathrm{d}\mathcal{H}^{d-1}(x)-\int_{A_s}\sigma(x)\,\mathrm{d}\mathcal{H}^{d-1}(x)\\ &\leq \big(\exp(C_\Phi(d-1)|s-t|)-1\big)\|\sigma\|_\infty\mathcal{H}^{d-1}(\hat{X})\\ &\quad+\exp\big(C_\Phi(d-1)|s-t|\big)\mathcal{H}^{d-1}(\hat{X})\omega(|s-t|/\varepsilon_{\mathrm{nd}})\\ &\leq C_3\cdot\big(|s-t|+\omega(|s-t|/\varepsilon_{\mathrm{nd}})\big),\end{aligned}$$

where the constant C_3 depends on C_L, $\varepsilon_{\mathrm{nd}}$, $\varepsilon_{\mathrm{tr}}$, $\|\sigma\|_\infty$ and $\operatorname{diam}(\hat{X})$. Exchanging the roles of s and t completes the proof of (B.7).

Proof of (B.8). By definition, for every point x in the set B_t, the curve $\Phi_i(x,[0,s-t])$ must cross the boundary of L at some point, so that

$$u(x):=\min\{v\in[0,s-t]\mid \Phi_i(x,v)\in\partial L\}$$

is well defined. We write $P(x):=\Phi_i(x,u(x))$ for the corresponding point on the boundary of L. By definition of $u(x)$, the curve $\Phi(x,[0,u(x)])$ is included in L, so that by Lemma B.2(i) we have $f_i(P(x))=t+u(x)$. This shows

$$P(B_t)\subseteq\Lambda:=\partial L\cap f_i^{-1}([t,s]). \tag{B.11}$$

We now prove that the map P satisfies a reverse-Lipschitz inequality. Note that for any point x in B_t,

$$x=\Phi_i(P(x),-u(x))=\Phi_i(P(x),t-f_i(P(x))).$$

Using the bounds (B.5) and (B.4), we find that for any x, y in B_t,

$$\begin{aligned}\|x-y\|&\leq\|\Phi_i(P(x),t-f_i(P(x)))-\Phi_i(P(y),t-f_i(P(y)))\|\\ &\leq\|\Phi_i(P(x),t-f_i(P(x)))-\Phi_i(P(y),t-f_i(P(x)))\|\\ &\quad+\|\Phi_i(P(y),t-f_i(P(x)))-\Phi_i(P(y),t-f_i(P(y)))\|\\ &\leq\exp(C_\Phi T_{\mathrm{tr}})\|P(x)-P(y)\|+|f_i(P(x))-f_i(P(y))|/\varepsilon_{\mathrm{nd}}\\ &\leq C'\|P(x)-P(y)\|,\end{aligned}$$

where $C':=\exp(C_\Phi)+C_L/\varepsilon_{\mathrm{nd}}$; we have used the fact that $T_{\mathrm{tr}}\leq 1$. We can now bound the $(d-1)$-Hausdorff measure of B_t in terms of that of Λ using this Lipschitz bound and the inclusion (B.11):

$$\mathcal{H}^{d-1}(B_t)\leq\mathcal{H}^{d-1}(P^{-1}(P(B_t)))\leq C'^{d-1}\mathcal{H}^{d-1}(\Lambda). \tag{B.12}$$

What remains to be done is to prove that the $(d-1)$-Hausdorff measure of Λ behaves like $\mathrm{O}(|s-t|)$, and this is where the transversality condition will enter.

Let us write

$$F_j:=\begin{cases}f_j^{-1}(\lambda_j), & j\neq 0,i,\\ \partial\hat{X}\cap\partial L, & j=0.\end{cases}$$

Then ∂L can be partitioned (up to an $\mathcal{H}^{d-1}$-negligible set) into faces $\partial L = \bigcup_{j\neq i}(F_j \cap L)$ and using the coarea formula on each of the facets we get (writing $B := f_i^{-1}([t,s])$)

$$\begin{aligned}\mathcal{H}^{d-1}(\Lambda) &= \sum_{j\neq i} \mathcal{H}^{d-1}(B\cap(F_j\cap L)) = \sum_{j\neq i}\int_{B\cap(F_j\cap L)} \mathrm{d}\mathcal{H}^{d-1}(x)\\ &= \sum_{j\neq i}\int_t^s \int_{S_u\cap(F_j\cap L)} \frac{1}{J_{ij}(x)}\,\mathrm{d}\mathcal{H}^{d-2}(x)\,\mathrm{d}u,\end{aligned} \tag{B.13}$$

where $J_{ij}(x)$ is the Jacobian of the restriction of f_i to the hypersurface F_j. More precisely,

$$J_{ij}(x) = \left\| \nabla f_i(x) - \left\langle \nabla f_i(x) \,\middle|\, \nabla f_j(x)\right\rangle \frac{\nabla f_j(x)}{\|\nabla f_j(x)\|^2}\right\| \quad \text{if } j \neq 0, i,$$

and

$$J_{i0}(x) = \|\nabla f_i(x) - \langle \nabla f_i(x) \mid v_0(x)\rangle v_0(x)\|,$$

where $v_0(x) \in \mathcal{N}_x\hat{X}$ is a unit vector. Since $\hat{X}$ is convex, for $\mathcal{H}^{d-1}$-a.e. $x \in \partial\hat{X}$ the normal cone $\mathcal{N}_x\hat{X}$ consists of only one direction, thus for such x there is a unique choice of $v_0(x)$. Let us write $v_i = \nabla f_i(x)/\|\nabla f_i(x)\|$ and v_j for either $\nabla f_j(x)/\|\nabla f_j(x)\|$ or $v_0(x)$; we then have, using (T),

$$\begin{aligned}J_{ij}(x)^2 &= \|\nabla f_i(x)\|^2\|v_i - \langle v_i \mid v_j\rangle v_j\|^2\\ &\geq \|\nabla f_i(x)\|^2(1-\langle v_i \mid v_j\rangle^2) \geq \varepsilon_{\mathrm{nd}}^2\varepsilon_{\mathrm{tr}}^2.\end{aligned} \tag{B.14}$$

Combining (B.13) and (B.14) gives

$$\begin{aligned}\mathcal{H}^{d-1}(\Lambda) &\leq \frac{1}{\varepsilon_{\mathrm{nd}}\varepsilon_{\mathrm{tr}}}\sum_{j\neq i}\int_t^s \mathcal{H}^{d-2}(S_u\cap(F_j\cap L))\,\mathrm{d}u\\ &= \frac{1}{\varepsilon_{\mathrm{nd}}\varepsilon_{\mathrm{tr}}}\int_t^s \mathcal{H}^{d-2}(S_u\cap\partial L)\,\mathrm{d}u.\end{aligned} \tag{B.15}$$

By definition, a point belongs to the intersection $S_u \cap \partial L$ if it lies in the singularity set $\Sigma(\boldsymbol{\lambda}(u))$, where $\boldsymbol{\lambda}(u) = (\lambda_1,\ldots,\lambda_{i-1},u,\lambda_{i+1},\ldots,\lambda_N)$. By Lemma 3.2,

$$\mathcal{H}^{d-2}(S_u\cap\partial L) \leq \mathcal{H}^{d-2}(\Sigma(\boldsymbol{\lambda}(u))) \leq C(d,\mathrm{diam}(\hat{X}))\frac{1}{\varepsilon_{\mathrm{tr}}}. \tag{B.16}$$

Combining (B.12), (B.15) and (B.16) we obtain $\mathcal{H}^d(B_t) \leq C|t-s|$, which implies (B.8) by the boundedness of σ. □

B.2. *Continuity of partial derivatives*

We prove that the function $\hat{G}$ defined in (3.1) is continuously differentiable by controlling the modulus of continuity of its partial derivatives given in (3.3). Again, we start with a slightly more general proposition.

Proposition B.3. *Let σ be a continuous function on $\hat{X}$ with modulus of continuity ω and $i \in \{1, \dots, N\}$. Consider the following function on the cube $Q := [-T_{\text{tr}}, T_{\text{tr}}]^N$:*

$$H(\boldsymbol{\lambda}) := \int_{K(\boldsymbol{\lambda}) \cap f_i^{-1}(\lambda_i)} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x).$$

Then H is uniformly continuous on Q with modulus of continuity

$$\omega_H(\delta) = C_1 \cdot (\omega(C_2\delta) + |\delta|),$$

where the constants only depend on $\|f_i\|_{\mathcal{C}^{1,1}(\hat{X})}$, $\operatorname{diam}(\hat{X})$, ε_{nd}, ε_{tr}, and $\|\sigma\|_\infty$.

Proof. Proposition B.1 implies that the function H is uniformly continuous with respect to changes of the ith variable. Let us now consider variations with respect to the jth variable with $j \neq i$ by introducing

$$h : [-T_{\text{tr}}, T_{\text{tr}}] \ni s \mapsto \int_{K(\lambda_1, \dots, \lambda_{j-1}, s, \lambda_{j+1}, \dots, \lambda_N) \cap f_i^{-1}(\lambda_i)} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x)$$

for some fixed $\boldsymbol{\lambda} \in [-T_{\text{tr}}, T_{\text{tr}}]^N$. We can rewrite the difference between two values of h using the coarea formula. As before, we assume $s > t$ to fix the signs and introduce $L' := \hat{X} \cap \bigcap_{k \notin \{i,j\}} K_k(\lambda_k)$ and $S := f_i^{-1}(\lambda_i)$. We have

$$\begin{aligned} h(s) - h(t) &= \int_{L' \cap K_j(s) \cap S} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x) - \int_{L' \cap K_j(t) \cap S} \sigma(x)\, \mathrm{d}\mathcal{H}^{d-1}(x) \\ &= \int_t^s \int_{L' \cap S \cap f_j^{-1}(u)} \frac{\sigma(x)}{J_{ij}(x)}\, \mathrm{d}\mathcal{H}^{d-2}(x)\, \mathrm{d}u, \end{aligned}$$

where the Jacobian factor J_{ij} is no less than $\varepsilon_{\text{nd}}\varepsilon_{\text{tr}}$ from (B.14). Therefore,

$$h(s) \le h(t) + \frac{\|\sigma\|_\infty}{\varepsilon_{\text{nd}}\varepsilon_{\text{tr}}} \int_t^s \mathcal{H}^{d-2}(L \cap S \cap f^{-1}(u))\, \mathrm{d}u. \tag{B.17}$$

Just as in the proof of Proposition B.1, the set $L \cap S \cap f^{-1}(u)$ is included in the set $\Sigma(\lambda_1, \dots, \lambda_{j-1}, u, \lambda_j, \dots, \lambda_N)$. Thus, by Lemma 3.2,

$$\mathcal{H}^{d-2}(L \cap S \cap f^{-1}(u)) \le \frac{C(d, \hat{X})}{\varepsilon_{\text{tr}}}. \tag{B.18}$$

Combining (B.17) and (B.18) we can see that the function h is Lipschitz with constant

$$C_h := C(d, \hat{X}) \frac{\|\sigma\|_\infty}{\varepsilon_{\text{nd}}\varepsilon_{\text{tr}}^2}.$$

Finally,

$$
\begin{aligned}
|H(\boldsymbol{\mu}) - H(\boldsymbol{\lambda})| &\leq \sum_{j=1}^{N} |H(\lambda_1, \ldots, \lambda_{k-1}, \mu_j, \ldots, \mu_N) - H(\lambda_1, \ldots, \lambda_j, \mu_{j+1}, \ldots, \mu_N)| \\
&\leq \omega_h(|\mu_i - \lambda_i|) + \sum_{j \neq i} C_h |\mu_j - \lambda_j| \\
&\leq \omega_h(\|\mu - \lambda\|_\infty) + (N-1) C_h \|\mu - \lambda\|_\infty,
\end{aligned}
$$

where ω_h is the modulus of continuity defined in Proposition B.1. This establishes the uniform continuity of the function H, with the desired modulus of continuity. □

B.3. Proof of Theorem 3.1

Proposition B.1 shows that the partial derivative $\hat{G}$ with respect to the variable λ_i exists and is given by (B.2). Applying Proposition B.3 with $\sigma(x) = \hat{\rho}(x)/\|\nabla f_i(x)\|$, we obtain $\mathcal{C}^{0,\alpha}$ regularity for each of the partial derivatives of $\hat{G}$ on the cube $Q := [-T_{\text{tr}}, T_{\text{tr}}]^N$ from the $\mathcal{C}^{0,\alpha}$ regularity of $\hat{\rho}$. Moreover, the $\mathcal{C}^{0,\alpha}$ constant of each partial derivative over Q is controlled by

$$
C(\operatorname{diam}(\hat{X}), \varepsilon_{\text{nd}}, \varepsilon_{\text{tr}}, \|\nabla f_i\|_{\mathcal{C}^{1,1}(X)}, \|\hat{\rho}\|_{\mathcal{C}^{0,\alpha}(X)}).
$$

Acknowledgments. JK acknowledges the support of National Science Foundation grant DMS-1700094. QM and BT would like to acknowledge the support of the French ANR through the grant ANR-16-CE40-0014 (MAGA). BT is also partially supported by LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).

References

[1] Cgal, Computational Geometry Algorithms Library. http://www.cgal.org

[2] Geograml, a programming library of geometric algorithms. http://alice.loria.fr/software/geogram/doc/html/

[3] Acosta, G., Durán, R. G.: An optimal Poincaré inequality in L^1 for convex domains. Proc. Amer. Math. Soc.**132**, 195–202 (2004) Zbl 1057.26010 MR 2021262

[4] Attouch, H., Buttazzo, G., Michaille, G.: Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization. SIAM (2014) Zbl 1311.49001 MR 3288271

[5] Aurenhammer, F., Hoffmann, F., Aronov, B.: Minkowski-type theorems and least-squares clustering. Algorithmica **20**, 61–76 (1998) Zbl 0895.68135 MR 1483422

[6] Benamou, J.-D., Froese, B. D., Oberman, A. M.: Numerical solution of the optimal transportation problem using the Monge–Ampère equation. J. Comput. Phys. **260**, 107–126 (2014) Zbl 1349.65554 MR 3151832

[7] Caffarelli, L. A.: The regularity of mappings with a convex potential. J. Amer. Math. Soc. **5**, 99–104 (1992) Zbl 0753.35031 MR 1124980

[8] Caffarelli, L. A., Kochengin, S. A., Oliker, V. I.: On the numerical solution of the problem of reflector design with given far-field scattering data. In: Monge Ampère Equation: Applications to Geometry and Optimization (Deerfield Beach, FL, 1997), Contemp. Math. 226, Amer. Math. Soc., 13–32 (1999) Zbl 0917.65104 MR 1660740

[9] Carlier, G., A. Galichon, Santambrogio, F.: From Knothe's transport to Brenier's map and a continuation method for optimal transport. SIAM J. Math. Anal. **41**, 2554–2576 (2010) Zbl 1203.49063 MR 2607321

[10] de Castro, P. M. M., Mérigot, Q., Thibert, B.: Intersection of paraboloids and application to Minkowski-type problems. In: Computational Geometry (SoCG'14), ACM, 308–317 (2014) Zbl 1397.68212 MR 3382311

[11] de Goes, F., Breeden, K., Ostromoukhov, V., Desbrun, M.: Blue noise through optimal transport. ACM Trans. Graphics **31**, no. 6, art. 171, 10 pp. (2012)

[12] de Goes, F., Wallez, C., Huang, J., Pavlov, D., Desbrun, M.: Power particles: an incompressible fluid solver based on power diagrams. ACM Trans. Graphics **34**, no. 4, art. 50, 11 pp. (2015) Zbl 1334.68276

[13] Figalli, A., Kim, Y.-H., McCann, R. J.: Hölder continuity and injectivity of optimal maps. Arch. Ration. Mech. Anal. **209**, 747–795 (2013) Zbl 1281.49037 MR 3067826

[14] Friedland, S., Nabben, R.: On Cheeger-type inequalities for weighted graphs. J. Graph Theory **41**, 1–17 (2002) Zbl 1005.05028 MR 1919163

[15] Gangbo, W., McCann, R. J.: The geometry of optimal transportation. Acta Math. **177**, 113–161 (1996) Zbl 0887.49017 MR 1440931

[16] Guillen, N., Kitagawa, J.: On the local geometry of maps with c-convex potentials. Calc. Var. Partial Differential Equations **52**, 345–387 (2015) Zbl 1309.35038 MR 3299185

[17] Householder, A. S.: The Theory of Matrices in Numerical Analysis. Dover Publ., New York (1975) Zbl 0329.65003 MR 0378371

[18] Hug, D.: Generalized curvature measures and singularities of sets with positive reach. Forum Math. **10**, 699–728 (1998) Zbl 0938.52004 MR 1652084

[19] Kim, Y.-H., Kitagawa, J.: On the degeneracy of optimal transportation. Comm. Partial Differential Equations **39**, 1329–1363 (2014) Zbl 1304.49093 MR 3208810

[20] Kim, Y.-H., McCann, R. J.: Continuity, curvature, and the general covariance of optimal transportation. J. Eur. Math. Soc. **12**, 1009–1040 (2010) Zbl 1191.49046 MR 2654086

[21] Kitagawa, J.: An iterative scheme for solving the optimal transportation problem. Calc. Var. Partial Differential Equations **51**, 243–263 (2014) Zbl 1297.49051 MR 3247388

[22] Lévy, B.: A numerical algorithm for L_2 semi-discrete optimal transport in 3D. ESAIM Math. Model. Numer. Anal. **49**, 1693–1715 (2015) Zbl 1331.49037 MR 3423272

[23] Loeper, G.: On the regularity of solutions of optimal transportation problems. Acta Math. **202**, 241–283 (2009)

[24] Loeper, G.: Regularity of optimal maps on the sphere: the quadratic cost and the reflector antenna. Arch. Ration. Mech. Anal. **199**, 269–289 (2011) Zbl 1231.35280 MR 2754343

[25] Loeper, G., Rapetti, F.: Numerical solution of the Monge–Ampère equation by a Newton's algorithm. C. R. Math. Acad. Sci. Paris **340**, 319–324 (2005) Zbl 1067.65119 MR 2754343

[26] Ma, X.-N., Trudinger, N. S., Wang, X.-J.: Regularity of potential functions of the optimal transportation problem. Arch. Ration. Mech. Anal. **177**, 151–183 (2005) Zbl 1072.49035 MR 2188047

[27] Mérigot, Q.: A multiscale approach to optimal transport. Computer Graphics Forum **30**, 1583–1592 (2011)

[28] Mirebeau, J.-M.: Discretization of the 3D Monge–Ampère operator, between wide stencils and power diagrams. ESAIM Math. Model. Numer. Anal. **49**, 1511–1523 (2015) Zbl 1330.65161 MR 3423234

[29] Oliker, V., Prussner, L.: On the numerical solution of the equation $\frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y}\right)^2 = f$ and its discretizations, I. Numer. Math. **54**, 271–293 (1989) Zbl 0659.65116 MR 0971703

[30] Rockafellar, R. T.: Convex Analysis. Princeton Math. Ser. 28, Princeton Univ. Press, Princeton, NJ (1970) Zbl 0193.18401 MR 0274683

[31] Saumier, L.-P., Agueh, M., Khouider, B.: An efficient numerical algorithm for the L^2 optimal transport problem with periodic densities. IMA J. Appl. Math. **80**, 135–157 (2015) Zbl 06409002 MR 3335154

[32] Schneider, R.: Convex Bodies: The Brunn–Minkowski Theory. Cambridge Univ. Press (1993) Zbl 0798.52001 MR 1216521

[33] Trudinger, N. S., Wang, X.-J.: On the second boundary value problem for Monge–Ampère type equations and optimal transportation. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (5) **8**, 143–174 (2009) Zbl 1182.35134 MR 2512204

[34] Villani, C.: Optimal Transport: Old and New. Springer (2009) Zbl 1156.53003 MR 2459454