Convergence of a Newton algorithm for semi-discrete optimal transport

Jun Kitagawa, Quentin Mérigot, Boris Thibert

CONVERGENCE OF A NEWTON ALGORITHM FOR SEMI-DISCRETE OPTIMAL TRANSPORT

JUN KITAGAWA, QUENTIN MÉRIGOT, AND BORIS THIBERT

Abstract. Many problems in geometric optics or convex geometry can be recast as optimal transport problems and a popular way to solve these 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's algorithm for this type of problems, 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.

Contents

1. Introduction

Many problems in geometric optics or convex geometry can be recast as optimal transport problems: this includes the far-field reflector problem, Alexandrov's curvature prescription problem, etc. A popular way to solve these problems numerically is to assume that the source probability measure is 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 [25, 6, 17] or algorithms that are much faster in practice but which come with no convergence guarantees [3, 23, 8, 18, 7]. Algorithms of the first kind rely on coordinate-wise 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's algorithm which is experimentally efficient and by proving the global convergence of this algorithm with optimal rates. The main assumptions 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 $(c_y)_{y\in Y}$ of functions on Ω. We assume that the functions (c_y) are of class $\mathcal{C}^{1,1}$ on Ω:

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

We consider a compact subset X of Ω and ρ a probability density on X, i.e. such that $\rho \mathrm{d}\mathcal{H}^d$ 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

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

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 leaving 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 is equal to the integral of ρ over its Voronoi cell,

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

If the price of bread is given by a function $\psi : Y \to \mathbb{R}$, customers leaving at location x in X make a compromise between walking cost and price by

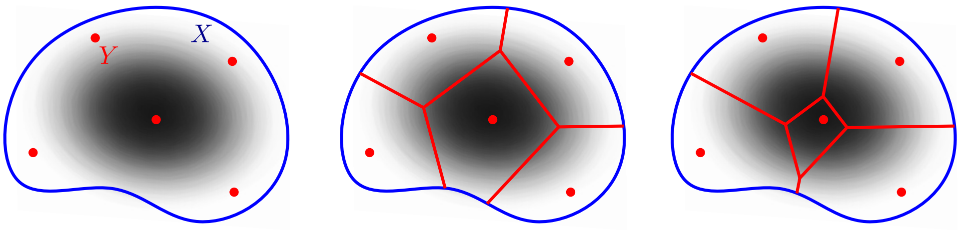

FIGURE 1. (Left) The domain X (with boundary in blue) is endowed with a probability density pictured in grayscale representing the density of population in a city. The set Y (in red) represents the location of bakeries. Here, $X, Y \subseteq \mathbb{R}^2$ and $c(x,y) = |x-y|^2$ (Middle) The Voronoi tessellation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

minimizing the sum $c(x,y) + \psi(y)$. This leads to the notion of Laguerre tessellation, whose cells are given by

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

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. The shape of the Voronoi and Laguerre tessellations is depicted in Figure 1.

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,

$$\forall x \in X,\ y \in Y \mapsto Dc_y(x) \in T^*_x\Omega \text{ is injective,} \tag{Twist}$$

where D is the differential of the function c_y. 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 (see Proposition 2.2), that T_ψ is an optimal transport map between ρ and the pushforward measure $T_{\psi\#}\rho = \sum_{y\in Y} \rho(\mathrm{Lag}_y(\psi))\delta_y$.

1.3. **Kantorovich's functional.** 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 [3] for the quadratic cost, but it is nonetheless included in Section 2 for the sake of completeness.

Here and after, we denote $(\mathbb{1}_y)_{y\in Y}$ the canonical basis of $\mathbb{R}^Y$, and $\|\cdot\|$ 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 of probability measures $\mathcal{P}(Y)$ as a subset of $\mathbb{R}^Y$.

THEOREM 1.1. *Assume* (Reg) *and* (Twist)*, let ρ be a bounded probability density on Ω and $\nu = \sum_{y\in Y} \nu_y \delta_y$ in $\mathcal{P}(Y)$. Then, the functional Φ*

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

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

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

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,*

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

We call *Kantorovich's functional* the function Φ introduced in (1.3). 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 Chapter 12 in [29]).

1.4. **Damped Newton algorithm.** We consider a simple damped Newton's algorithm to solve semi-discrete optimal transport problem. This algorithm is very close to the one used in [24] and has been suggested to us by Mirebeau. 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

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

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

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.

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

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

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

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

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

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

Algorithm 1: Simple damped Newton's algorithm

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 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 [22], which appeared naturally in the study of the regularity of optimal transport maps. We use a suitable discretization of Loeper's geometric reformulation of the Ma-Trundinger-Wang condition [19].

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

$$p \in \Omega_y \mapsto c(\exp_y^c p, y) - c(\exp_y^c p, z) \tag{QC}$$

are quasi-convex for all z in Y. The map $\exp_y^c$ is called the *c-exponential* with respect to y, and the domain Ω_y is an *exponential chart.*

Definition 1.2 (c-Convexity)**.** Assuming Loeper's condition, a subset X of Ω is *c-convex* if for every y in Y, the inverse image $(\exp_y^c)^{-1}(X)$ is convex.

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 Laguerre cells in the exponential charts, which 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 in $\mathcal{P}^{\mathrm{ac}}(X) \cap \mathcal{C}^\alpha(X)$ for α in* $(0, 1]$*. Then, Kantorovich's functional is of class $\mathcal{C}^{2,\alpha}_{loc}$ 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

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

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 [22], which include the well known (MTW) or (A3) condition. See Remark 4.3 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 [22, 28], there are the nonflat examples of Riemannian distance squared and $-\log\|x-y\|_{\mathbb{R}^{n+1}}$ on (a subset of) $\mathbb{S}^n \times \mathbb{S}^n$ (see [20]). The last cost is associated to the *far-field reflector antenna problem.* We refer the reader to [15, 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,

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

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

We denote 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), (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 assuming 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 $\rho \in \mathcal{C}^\alpha(X)$ for α in* $(0, 1]$.
(ii) *ρ has positive Poincaré-Wirtinger* (PW) *constant.*

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

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 [23, 8, 4]. Note however that none of these works contain convergence proofs for the Newton's algorithm.

Loeper and Rappetti [21] (refined by Saumier, Agueh, and Khouider [26]) establish the global convergence of a damped Newton's 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 construct explicitly probability densities with non-convex support that still satisfy the hypothesis of Theorem 1.5). A second drawback on relying on the regularity theory for optimal transport is that the damping parameter, which is an input parameter of the algorithm used in [21], 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 [21][26] can be adapted to cost functions satisfying the Ma-Trudinger-Wang condition (which is equivalent to Loeper's condition (QC) that we also require).

Finally Oliker and Prussner prove the *local* convergence of Newton's method for finding Alexandrov's solutions to the Monge-Ampère equation $\det \mathrm{D}^2 u = \nu$ with Dirichlet boundary conditions, where ν is a finitely supported measure [25]. Global convergence for a damped Newton's algorithm is established by Mirebeau [24] for a variant of Oliker and Prussner's scheme, but without convergence rates. Theorem 1.5 article can be seen as an extension of the strategy used 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.

Outline. In Section 2, we establish the differentiability of Kantorovich's functional Φ, adapting arguments from [3]. 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's 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 an elementary proof of a convex geometry result used in the regularity proof. Finally, Appendix C contains an alternate proof of a crucial transversality condition, under the assumption that the target set Y is sampled from some continuous domain.

2. Kantorovich's functional

The purpose of this section is to present the variational formulation introduced in [3] for the semi-discrete optimal transport problem, adapting the arguments presented for the squared Euclidean cost in [3] to cost functions satisfying (Reg') and (Twist'), which are weaker than the conditions (Reg) and (Twist) presented in the introduction:

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

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

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, we include proofs for completeness.

Theorem 2.1. *Assume* (Reg') *and* (Twist')*, and let ρ be a bounded probability density on Ω and $\nu = \sum_{y\in Y} \nu_y \delta_y$ be 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 pushforward 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 one has

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

Multiplying this inequality by ρ and integrating it over X gives

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

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

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

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

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

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:*

$$G(\psi) = \left(\rho(\mathrm{Lag}_y(\psi))\right)_{y\in Y} \tag{2.9}$$

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, $g(t) - h(t) = \rho(f^{-1}(t)) = 0$. Moreover, using Lebesgue's monotone convergence theorem one easily sees that g (resp. h) is right-continuous (resp. left-continuous). This concludes the proof $\square$

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 remark that by definition, $\mathrm{Lag}_y(\psi) = \bigcap_{z\neq y\in Y} H_z(\psi)$ where

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

Denoting $A\Delta B$ the symmetric difference of two sets, we have the following inequalities

$$|G_y(\psi) - G_y(\varphi)| \leqslant \rho(\mathrm{Lag}_y(\psi)\Delta \mathrm{Lag}_y(\varphi)) \leqslant \sum_{z\neq y\in Y} \rho(H_z(\psi)\Delta H_z(\varphi)). \tag{2.10}$$

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

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

By (Twist') and Lemma 2.4 we know that $\lim_{\varphi\to\psi} \rho(H_z(\psi)\Delta H_z(\varphi)) = 0$, which with (2.10) concludes the proof. $\square$

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 $\min_{y\in Y}(c(x, y) + \psi(y)) \leqslant c(y, T(y)) + \psi(T(y))$, which by integration gives

$$\Phi(\psi) \leqslant \int_X (c(x, T(x)) + \psi(T(x)))\rho(x)\mathrm{d}\mathcal{H}^d_g(x) - \sum_{y\in Y} \psi(y)\nu_y. \tag{2.11}$$

Moreover, equality holds when $T = T_\psi$. Taking another function φ on Y and setting $T = T_\varphi$ in Equation (2.11) gives

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

where G is defined as in the statement of Proposition 2.3. This proves that the superdifferential of Φ is never empty, thus establishing the concavity. By Proposition 2.3, the map G is continuous, meaning that we have constructed a continuous selection of a supergradient in the superdifferential of the concave function Φ. By standard arguments from convex analysis, this proves that Φ is $\mathcal{C}^1$, and that $\nabla\Phi(\psi) = G(\psi) - \nu$.

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 $\kappa \in \mathbb{R}$ the closed sublevel sets $K_i(\kappa) := f_i^{-1}([-\infty, \kappa])$ are convex. Let $\hat{\rho}$ be a continuous probability density over $\hat{X}$. The purpose of this section is to give sufficient conditions to ensure the regularity of the following function $\hat{G}$ near the origin of $\mathbb{R}^N$:

$$\hat{G} : \boldsymbol{\lambda} \in \mathbb{R}^N \mapsto \int_{K(\boldsymbol{\lambda})} \hat{\rho}(x) \mathrm{d}\mathcal{H}^d(x), \tag{3.12}$$

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

3.1. Assumptions and statement of the theorem.

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

Non-degeneracy. The functions (f_i) satisfy the *nondegeneracy condition* if the norm of their gradients is bounded from below:

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

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 $(K(\boldsymbol{\lambda}) \cap \partial K_i(\lambda_i))_{1 \leqslant i \leqslant N}$ and $K(\boldsymbol{\lambda}) \cap \partial \hat{X}$. The purpose of the transversality condition we consider is to ensure that adjacent facets make a minimum angle, when $\boldsymbol{\lambda}$ remains close to some fixed vector $\boldsymbol{\lambda}_0$. More precisely, we say that the family of functions (f_i) satisfy the *transversality condition near* $\boldsymbol{\lambda}_0$ if there exists positive constants ε_{tr} and $T_{tr} \leqslant 1$ such that for every $\boldsymbol{\lambda}$ in $\mathbb{R}^N$ satisfying $\|\boldsymbol{\lambda} - \boldsymbol{\lambda_0}\|_\infty \leqslant T_{tr}$ for the usual ℓ^∞ norm on $\mathbb{R}^N$ and every point x in $\partial K(\boldsymbol{\lambda})$ one has,

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

where $\mathcal{N}_x \hat{X}$ is the normal cone to the convex set $\hat{X}$ at x (see (B.68) for a definition). When $\partial \hat{X}$ is smooth at x, $\mathcal{N}_x \hat{X}$ is the ray spanned by the exterior normal to $\hat{X}$ at x. We denote by $\Sigma(\boldsymbol{\lambda})$ the set of points x in the boundary of $K(\boldsymbol{\lambda})$ satisfying one of the assumptions in (T) or, equivalently, which belong to two (or more) distinct facets of $K(\boldsymbol{\lambda})$:

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

We will see 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 $\Sigma(\boldsymbol{\lambda})$, which is essential in establishing the smoothness of $\hat{G}$.

THEOREM 3.1. *Assume that the functions satisfy the non-degeneracy condition* (ND) *and the transversality condition* (T) *near* $\boldsymbol{\lambda_0}$. *Let* $\hat{\rho}$ *be a* $\mathcal{C}^\alpha$ *probability density on* $\hat{X}$. *Then, the map* $\hat{G}$ *defined in* (3.12) *is of class* $\mathcal{C}^{1,\alpha}$ *on the cube* $Q := \boldsymbol{\lambda_0} + [-T_{tr}, T_{tr}]^N$ *and has partial derivatives given by*

$$\frac{\partial \hat{G}}{\partial \lambda_i}(\boldsymbol{\lambda}) = \int_{K(\boldsymbol{\lambda})\cap\partial K_i(\boldsymbol{\lambda})} \frac{\hat{\rho}(x)}{\|\nabla f_i(x)\|} \mathrm{d}\mathcal{H}^{d-1}(x). \tag{3.13}$$

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

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

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

3.2. **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 for the map $\hat{G}$. In this section, we denote $e_1, \dots, e_N$ 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 $\|\boldsymbol{\lambda}\| < T_{tr}$. For $t > 0$, one has:

$$\begin{aligned}\frac{1}{t}(\hat{G}(\boldsymbol{\lambda} + t\boldsymbol{e}_i) - \hat{G}(\boldsymbol{\lambda})) &= \frac{1}{t}\int_{K(\boldsymbol{\lambda}+t\boldsymbol{e}_i)\setminus K(\boldsymbol{\lambda})} \hat{\rho}(x)\mathrm{d}\mathcal{H}^d(x)\\ &= \frac{1}{t}\int_{\lambda_i}^{\lambda_i+t} \hat{g}(s)\mathrm{d}s,\end{aligned} \tag{3.14}$$

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

$$\hat{g}(s) := \int_{\cap_{j\neq i}K_j(\lambda_j)\cap f_i^{-1}(s)} \frac{\hat{\rho}(x)}{\|\nabla f_i(x)\|}\mathrm{d}\mathcal{H}^{d-1}(x). \tag{3.15}$$

The same reasoning also holds for $t < 0$. We now claim that $\hat{g}$ is continuous on some interval around λ_i, which by (3.14) and the Fundamental Theorem of Calculus will imply that the limit as $t \to 0$ in (3.14) exists and is equal to $\hat{g}(\boldsymbol{\lambda})$, thus establishing the formula (3.13). The continuity of $\hat{g}$ follows from the next proposition, which is formulated in a slightly more general way.

PROPOSITION 3.2. *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*

$\|\boldsymbol{\lambda}\|_\infty \leqslant T_{tr}$, *consider the function*

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

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

$$\omega_h(\delta) = \mathrm{const}\cdot(\omega(C\delta) + |\delta|), \tag{3.16}$$

where the constants only depend on $\|f_i\|_{\mathcal{C}^{1,1}}$, $\mathrm{diam}(\hat{X})$, ε_{nd}, ε_{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 (3.15) is continuous. This implies the existence of partial derivatives and establishes formula (3.13). The proof of Proposition 3.2 requires two lemmas.

LEMMA 3.3. *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) The map Φ_i *satisfies the following inequalities for every* $x, y \in \hat{X}$, $t \in \mathbb{R}$*:*

$$\|\Phi_i(x, t) - \Phi_i(x, s)\| \leqslant \frac{|t - s|}{\varepsilon_{nd}} \tag{3.17}$$

$$\|\Phi_i(x, t) - \Phi_i(y, t)\| \leqslant \exp(C_\Phi |t|)\,\|x - y\|, \tag{3.18}$$

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

Proof. We consider the vector field $V_i^0(x) = \nabla f_i(x)/\|\nabla f_i(x)\|^2$ on $\hat{X}$, which satisfies $\|V_i^0\|_\infty \leqslant 1/\varepsilon_{nd}$ 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 $p_{\hat{X}}$,

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

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 and is for all time since V_i is bounded and uniformly Lipschitz on all of $\mathbb{R}^d$. The inequality (3.17) 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 (3.18) follows from the bound on the Lipschitz constant of V_i and from Gronwall's lemma. □

LEMMA 3.4. *Assuming the transversality condition* (T), *there exists a constant depending only on* d *and* $\mathrm{diam}(\hat{X})$ *such that for every* $\|\boldsymbol{\lambda}\|_\infty \leqslant T_{tr}$,

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

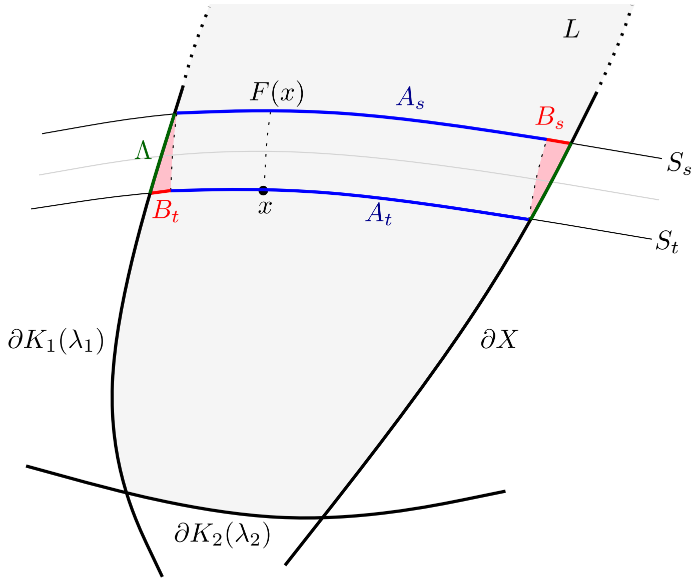

FIGURE 2. Illustration of the proof of Proposition 3.2.

Proof. Given $\|\boldsymbol{\lambda}\|_\infty \leqslant T_{tr}$, the transversality condition (T) implies

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

where $\mathcal{N}_x K(\boldsymbol{\lambda})$ is the normal cone to the convex set $K(\boldsymbol{\lambda})$ at x (see (B.68)). This implies that $\Sigma(\boldsymbol{\lambda})$ is included in the set $\mathrm{Sing}(K(\boldsymbol{\lambda}), \varepsilon_{tr})$ of τ-singular points defined in (B.69), with $\tau = \varepsilon_{tr}$. The lemma then follows from Proposition B.1. $\square$

Proof of Proposition 3.2. Let t, s be small enough so that the transversality condition (T) holds (that is $t, s \in [-T_{tr}, T_{tr}]$). We assume that $t < s$ so as 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 2:

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

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,

$$h(t) = \int_{A_t} \sigma(x) \mathrm{d}\mathcal{H}^{d-1}(x) + \int_{B_t} \sigma(x) \mathrm{d}\mathcal{H}^{d-1}(x), \tag{3.19}$$

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 term in the sums defining $h(t)$ and $h(s)$ in (3.19) are close, namely

$$\left|\int_{A_t} \sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x) - \int_{A_s} \sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x)\right| \leqslant \mathrm{const}\cdot(|s-t| + \omega(C\,|s-t|)) \tag{3.20}$$

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

$$\int_{B_t} \left|\sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x)\right| + \int_{B_s} \left|\sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x)\right| \leqslant \mathrm{const}\cdot|s-t| \tag{3.21}$$

The combination of the estimates (3.20) and (3.21) implies the desired inequality (3.16). We now turn to the proof of these estimates, and that the constant in these estimates depend on $\|f_i\|_{\mathcal{C}^{1,1}}$, $\mathrm{diam}(\hat{X})$, ε_{nd}, ε_{tr} and $\|\sigma\|_\infty$.

Proof of Estimate (3.20)**.** By Lemma 3.3.(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 (3.18), the restriction of F to A_t is a bi-Lipschitz bijection between the sets A_t and A_s, with Lipschitz constant

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

Using a Lipschitz change of variable formula, we get

$$\begin{aligned}\int_{A_t} \sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x) &= \int_{F^{-1}(A_s)} \sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x)\\ &\leqslant \left\|F^{-1}\right\|_{\mathrm{Lip}(A_s)}^{d-1} \int_{A_s} \sigma(F^{-1}(x))\mathrm{d}\mathcal{H}^{d-1}(x)\\ &\leqslant \exp(C_\Phi(d-1)\,|s-t|) \int_{A_s} \sigma(F^{-1}(x))\mathrm{d}\mathcal{H}^{d-1}(x).\end{aligned} \tag{3.22}$$

By definition of the modulus of continuity and thanks to (3.17),

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

Integrating this inequality, we get

$$\begin{aligned}\int_{A_s} \sigma(F^{-1}(x))\mathrm{d}\mathcal{H}^{d-1}(x) &\leqslant \int_{A_s} \sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x) + \mathcal{H}^{d-1}(A_s)\omega(|s-t|\,/\varepsilon_{nd})\\ &\leqslant \int_{A_s} \sigma(x)\mathrm{d}\mathcal{H}^{d-1}(x) + \mathcal{H}^{d-1}(\hat{X})\omega(|s-t|\,/\varepsilon_{nd})\end{aligned} \tag{3.23}$$

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 [27, p.211]. Combining (3.22) and (3.23) we get

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

so that

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

where the constant depends on C_L, ε_{nd}, ε_{tr}, $\|\sigma\|_\infty$ and $\mathrm{diam}(\hat{X})$. Exchanging the role of s and t completes the proof of (3.20).

Proof of (3.21)**.** 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 3.3.(i) we have $f_i(P(x)) = t + u(x)$. This shows

$$P(B_t) \subseteq \Lambda := \partial L \cap f_i^{-1}([t, s]). \tag{3.24}$$

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 (3.18) and (3.17), we get that for any x, y in B_t,

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

where $C' := \exp(C_\Phi) + C_L/\varepsilon_{nd}$; we have used that $T_{tr} \leqslant 1$. We can now bound the $(d-1)$–Hausdorff measure of B_t from that of Λ using this Lipschitz bound and the inclusion (3.24):

$$\mathcal{H}^{d-1}(B_t) \leqslant \mathcal{H}^{d-1}(P^{-1}(P(B_t))) \leqslant C'^{d-1}\mathcal{H}^{d-1}(\Lambda). \tag{3.25}$$

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 a $\mathcal{H}^{d-1}$–negligible set) into faces $\partial L = \cup_{j\neq i}(F_j \cap L)$ and using the coarea formula on each of the facets we get (writing $B := f_i^{-1}([t,s])$)

$$\begin{aligned}\mathcal{H}^{d-1}(\Lambda) &= \sum_{j\neq i} \mathcal{H}^{d-1}(B\cap(F_j\cap L))\\ &= \sum_{j\neq i}\int_{B\cap(F_j\cap L)} \mathrm{d}\mathcal{H}^{d-1}(x)\\ &= \sum_{j\neq i}\int_t^s\int_{S_u\cap(F_j\cap L)} \frac{1}{J_{ij}(x)}\mathrm{d}\mathcal{H}^{d-2}(x)\mathrm{d}u, \end{aligned}\tag{3.26}$$

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

$$J_{ij}(x) = \left\| \nabla f_i(x) - \langle \nabla f_i(x) \mid \nabla f_j(x)\rangle \frac{\nabla f_j(x)}{\|\nabla f_j(x)\|^2}\right\|$$

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)

$$\begin{aligned} J_{ij}(x)^2 &= \|\nabla f_i(x)\|^2\|v_i - \langle v_i\mid v_j\rangle v_j\|^2\\ &\geqslant \|\nabla f_i(x)\|^2(1-\langle v_i\mid v_j\rangle^2)\\ &\geqslant \varepsilon_{nd}^2\varepsilon_{tr}^2. \end{aligned}\tag{3.27}$$

Combining (3.26) and (3.27) gives us

$$\begin{aligned}\mathcal{H}^{d-1}(\Lambda) &\leqslant \frac{1}{\varepsilon_{nd}\varepsilon_{tr}}\sum_{j\neq i}\int_t^s \mathcal{H}^{d-2}(S_u\cap(F_j\cap L))\mathrm{d}u\\ &= \frac{1}{\varepsilon_{nd}\varepsilon_{tr}}\int_t^s \mathcal{H}^{d-2}(S_u\cap\partial L)\mathrm{d}u. \end{aligned}\tag{3.28}$$

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 $\boldsymbol{\lambda}(u) = (\lambda_1,\dots,\lambda_{i-1},u,\lambda_{i+1},\dots,\lambda_N)$. By Lemma 3.4,

$$\mathcal{H}^{d-2}(S_u\cap\partial L)\leqslant \mathcal{H}^{d-2}(\Sigma(\boldsymbol{\lambda}(u)))\leqslant \mathrm{const}(d,\mathrm{diam}(\hat{X}))\cdot\frac{1}{\varepsilon_{tr}}.\tag{3.29}$$

Combining (3.25), (3.28) and (3.29) we obtain $\mathcal{H}^d(B_t)\leqslant \mathrm{const}\cdot|t-s|$, which implies (3.21) using the boundedness of σ. □

3.3. **Continuity of partial derivatives.** We prove that the function $\hat{G}$ defined in (3.12) is continuously differentiable by controlling the modulus of continuity of its partial derivatives given in (3.13). Again, we start with a slightly more general proposition.

PROPOSITION 3.5. *Let σ be a continuous function on $\hat{X}$ with modulus of continuity ω. Consider the following function on the cube $Q := [-T_{tr}, T_{tr}]^N$:*

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

Then H is uniformly continuous on Q with modulus of continuity

$$\omega_H(\delta) = \mathrm{const}\cdot(\omega(C\delta) + |\delta|),$$

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

Proof. Proposition 3.2 yields 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

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

for some fixed $\boldsymbol{\lambda} \in [-T_{tr}, T_{tr}]^N$. 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 $J_{ij} \geqslant \varepsilon_{nd}\varepsilon_{tr}$ from (3.27). Therefore,

$$h(s) \leqslant h(t) + \frac{\|\sigma\|_\infty}{\varepsilon_{nd}\varepsilon_{tr}} \int_t^s \mathcal{H}^{d-2}(L\cap S\cap f^{-1}(u))\mathrm{d}u. \tag{3.30}$$

Just as in the proof of Proposition 3.2, 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 3.4,

$$\mathcal{H}^{d-2}(L\cap S\cap f^{-1}(u)) \leqslant \frac{\mathrm{const}(d,\hat{X})}{\varepsilon_{tr}}. \tag{3.31}$$

Combining (3.30) and (3.31) we can see that the function h is Lipschitz with constant

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

Finally,

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

where ω_h is the modulus of continuity defined in Proposition 3.2. This establishes the uniform continuity of the function H, with the desired modulus of continuity. □

3.4. **Proof of Theorem 3.1.** Proposition 3.2 shows that the partial derivative $\hat{G}$ with respect to the variable λ_i exists and that its expression is given by (3.15). Applying Proposition 3.5 with $\sigma(x) = \hat{\rho}(x)/\left\|\nabla f_i(x)\right\|$, we obtain $\mathcal{C}^\alpha$ regularity for each of the partial derivatives of $\hat{G}$ on the cube $Q := [-T_{tr}, T_{tr}]^N$ from the $\mathcal{C}^\alpha$ regularity of $\hat{\rho}$. Moreover, the $\mathcal{C}^\alpha$ constant of each partial derivative over Q is controlled by

$$\mathrm{const}(\mathrm{diam}(\hat{X}), \varepsilon_{nd}, \varepsilon_{tr}, \left\|\nabla f_i\right\|_{\mathcal{C}^{1,1}(X)}, \left\|\hat{\rho}\right\|_{\mathcal{C}^\alpha(X)}).$$

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

This section is devoted to the proof of the following regularity result.

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

$$\mathcal{K}^\varepsilon := \{\psi : Y \to \mathbb{R} \mid \forall y \in Y,\ \rho(\mathrm{Lag}_y(\psi)) > \varepsilon\} \tag{4.32}$$

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 $X_y = (\exp_y^c)^{-1}(X) \subseteq \mathbb{R}^d$ 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

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

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

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

$$\begin{aligned}
\varepsilon_{tw} &:= \min_{x\in X} \min_{\substack{y,z\in Y\\ y\neq z}} \left\|Dc_y(x) - Dc_z(x)\right\|_g > 0\\
C_\nabla &:= \max_{(x,y)\in X\times Y} \left\|Dc_y(x)\right\|_g < +\infty\\
C_{\exp} &:= \max_{y\in Y} \max\left\{\left\|\exp_y^c\right\|_{\mathrm{Lip}(X_y)}, \left\|(\exp_y^c)^{-1}\right\|_{\mathrm{Lip}(X)}\right\} < +\infty,
\end{aligned} \tag{4.33}$$

where we recall that $c_y(x) = c(x, y)$ and $X_y := \exp_y^{-1}(X)$. 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:

$$C_{\mathrm{cond}} := \max_{y\in Y} \max_{p\in X_y} \mathrm{cond}(\mathrm{D}\exp_y^c\big|_p),$$
$$C_{\det} := \max_{y\in Y} \left\|\det(\mathrm{D}\exp_y^c)\right\|_{\mathrm{Lip}(X_y)},$$

where $\mathrm{cond}(A)$ is the condition number of a linear transform A between finite dimensional normed spaces and $\det(A)$ is the determinant of A with respect to orthornormal bases. The quantitative transversality estimates involve all the above constants in an explicit way, see (4.45).

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 $y_\pm = \pm 1$, let $y_0 = 0$, $Y = \{y_-, y_0, y_+\} \subseteq \mathbb{R}$. Consider the cost $c(x, y) = -xy$, and the density $\rho = 1$ on $X = [-\frac{1}{2}, \frac{1}{2}]$. 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 $\tau \geqslant 0$,

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

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

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

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 (B.68).

PROPOSITION 4.2. *Assume that ψ lies in $\mathcal{K}^{\varepsilon/2}$ (see (4.32)). 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*

$$\langle v \mid w \rangle \geqslant -1 + \delta_0^2, \tag{4.34}$$

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

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 3). 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 (4.34).

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

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

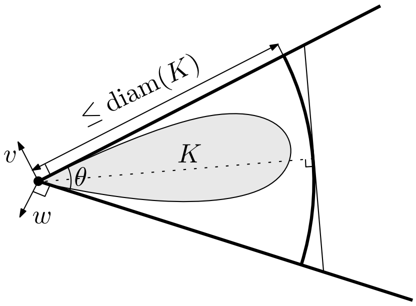

FIGURE 3. Bound on the volume of a convex set K as a function of the angle between two normal vectors v, w at the same point and its diameter (see Lemma 4.3).

Proof. The left-hand side of the inequality is non-positive, so the inequality needs only to be proven when $\langle v \mid w\rangle \leqslant 0$, which we assume from now on. Making a rotation of axes and a translation if necessary, we assume that p lies at the orign and that the unit vectors span the first two coordinates of $\mathbb{R}^d$. Then, letting $H := \{p \mid \langle p \mid v\rangle \leqslant 0\}$, $H' := \{p \mid \langle p \mid w\rangle \leqslant 0\}$ 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 3). Therefore, we have

$$\begin{aligned}\mathcal{H}^d(K) &\leqslant \mathcal{H}^d(H \cap H' \cap (D \times [-\operatorname{diam}(K), \operatorname{diam}(K)]^{d-2})) \\ &\leqslant 2^{d-2} \operatorname{diam}(K)^d \tan(\theta/2). \end{aligned} \tag{4.35}$$

Using the expression of $\cos(\theta)$ in term of $\tan(\theta/2)$ and recalling $\langle v \mid w\rangle \leqslant 0$,

$$\tan(\theta/2) = \sqrt{\frac{1 + \langle v \mid w\rangle}{1 - \langle v \mid w\rangle}} \leqslant \sqrt{1 + \langle v \mid w\rangle} \tag{4.36}$$

The lemma follows directly from Equations (4.35)–(4.36). □

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

$$\mathcal{H}^d(L_y(\psi)) \geqslant \varepsilon/(2C_{\exp}^d \left\|\rho\right\|_\infty) \text{ and } \operatorname{diam}(L_y(\psi)) \leqslant C_{\exp} \operatorname{diam}(X).$$

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

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

We also record the following lemma for later use.

LEMMA 4.4. *Let y in Y and let p in 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 $z' \in Y$,

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

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

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 differential of exponential maps, see Remark 4.1. The advantage of this first 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 in throughout this section. We fix a point y_0 in Y and also fix an arbitrary ordering of the remaining points, so that $Y = \{y_0, y_1, \ldots, y_N\}$. We define $\hat{X} := X_{y_0}$ and for every index $i \in \{1, \ldots, N\}$ we put

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

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

$$\varepsilon_{nd} := \min_{i,j\neq 0} \min_{p\in X_{y_0}} \|\nabla f_i(p) - \nabla f_j(p)\| \geqslant C_{\exp}^{-1}\varepsilon_{tw} > 0, \tag{4.37}$$

$$\sup_{i\neq 0} \sup_{p\in X_{y_0}} \|\nabla f_i(p)\| \leqslant C_{\exp} C_\nabla. \tag{4.38}$$

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

$$\boldsymbol{\lambda}_\psi : = (\psi(y_1) - \psi(y_0), \ldots, \psi(y_N) - \psi(y_0)) \in \mathbb{R}^N. \tag{4.39}$$

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

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

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

Proposition 4.5. *Assume that $\boldsymbol{\lambda} := \boldsymbol{\lambda}_\psi$ where ψ belongs to $\mathcal{K}^{\varepsilon/2}$ and let p be a point in $K(\boldsymbol{\lambda})$. Then,*

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

$$\left(\frac{\langle \nabla f_i(p) \mid \nabla f_j(p)\rangle}{\|\nabla f_i(p)\| \, \|\nabla f_j(p)\|}\right)^2 \leqslant 1 - \delta_1^2. \tag{4.40}$$

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

$$\forall w \in \mathcal{N}_p\hat{X}, \ \left(\frac{\langle \nabla f_i(p) \mid w\rangle}{\|\nabla f_i(p)\| \, \|w\|}\right)^2 \leqslant 1 - \delta_1^2. \tag{4.41}$$

In the above inequalities,

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

By assumption, all the Laguerre cells associated to ψ contain a mass of at least $\varepsilon/2$. This allows us to apply Proposition 4.2, ensuring that normal vectors cannot be near-opposite, to all the Laguerre cells in their exponential charts. 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, 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 one has*

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

where A^* denotes the adjoint of A.

Proof. Consider x in ∂K, $v \in \mathcal{N}_x K$, 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 the set $F(K)$ is non-positive. Since $F(K)$ is convex, for any point $y \in K$, $F(K)$ contains the segment $[F(x), F(y)]$. We therefore have

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

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

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

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 image under a linear map, using the generalized Wiedlandt inequality (see [13, 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 be a point in X and set $p_k := (\exp^c_{y_k})^{-1}(x)$ and $p_\ell := (\exp^c_{y_\ell})^{-1}(x)$ 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, the following inequalities hold for all v, w in $\mathbb{R}^d$:

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

Proof. Indeed, let θ be the angle between v and w and θ' be the angle between Av and Aw, both in the interval $(0, \pi)$. Let $t := \tan(\theta/2)$, $t' :=$

$\tan(\theta'/2)$. The generalized Wiedlandt inequality in [13, Section 3.4] asserts $(1/\operatorname{cond}(A))t \leqslant t' \leqslant \operatorname{cond}(A)t$. Expressing $\cos(\theta)$ in term of $t = \tan(\theta/2)$,

$$1 + \cos(\theta') = 1 + \frac{1 - t'^2}{1 + t'^2} = \frac{2}{1 + t'^2} \leqslant \operatorname{cond}(A)^2(1 + \cos(\theta))$$

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

Proof of Proposition 4.5, case I. We let

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

Switching the indices i and j if necessary, we assume that $\|V\| \leqslant \|W\|$. The proof depends on the sign of $\langle W - V \mid V\rangle$. Assume first $\langle W - V \mid V\rangle \leqslant 0$, and let $\alpha_v := 1/\|V\|, \alpha_w := 1/\|W\|$. Then,

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

Using $\alpha_w \geqslant \alpha_v$, and $\|W - V\| \geqslant \varepsilon_{nd}$ we end up with

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

where we have used (4.37) and (4.38), $\delta_0 \leqslant 1$ and $C_{\text{cond}} \geqslant 1$. This establishes the desired bound when $\langle v \mid w\rangle \in [0, 1]$. In the case $\langle v \mid w\rangle \in [-1, 0]$, we can apply Proposition 4.2 to show that $1 - \langle v \mid w\rangle^2 \geqslant 1 + \langle v \mid w\rangle \geqslant \delta_0^2 \geqslant \delta_1^2$, thus establishing the desired bound.

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

$$\begin{aligned} \langle v \mid w\rangle^2 &= 1 - \frac{\|W - V\|^2}{\|W\|^2} + \frac{\langle W - V \mid v\rangle^2}{\|W\|^2} \\ &= 1 - \frac{\|W - V\|^2}{\|W\|^2}\left(1 - \frac{\langle W - V \mid v\rangle^2}{\|W - V\|^2}\right) \\ &\leqslant 1 - \frac{\varepsilon_{nd}^2}{C_{\exp}^2 C_\nabla^2}\left(1 - \langle \frac{W - V}{\|W - V\|} \mid v\rangle\right), \end{aligned} \tag{4.42}$$

where we have used $\langle W - V \mid V\rangle \geqslant 0$ with (4.37) and (4.38) to get the last inequality. We will now apply Proposition 4.2 to the Laguerre cell L_i. By Lemma 4.4, the point $p_i := (\exp_{y_i}^c)^{-1}(\exp_{y_0}^c(p)) \in X_{y_i}$ belong 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 normals to L_i at p_i. Proposition 4.2 then shows that the vectors V_i and W_i satisfy

$$-1 + \delta_0^2 \leqslant \frac{\langle V_i \mid W_i\rangle}{\|V_i\|\,\|W_i\|}. \tag{4.43}$$

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

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

First, note that $W - V = AW_i$ and $V = -AV_i$. Applying the generalized Wiedlandt inequality (Lemma 4.7) and (4.43) we have

$$\begin{aligned} 1 - \frac{\langle W - V \mid v\rangle}{\|W - V\|} = 1 + \frac{\langle AW_i \mid AV_i\rangle}{\|AW_i\|\,\|AV_i\|} &\geqslant C_{\mathrm{cond}}^{-4}\left(1 + \frac{\langle V_i \mid W_i\rangle}{\|V_i\|\,\|W_i\|}\right) \\ &\geqslant C_{\mathrm{cond}}^{-4}\delta_0^2 \geqslant \delta_1^2. \end{aligned} \tag{4.44}$$

Combining this inequality with (4.42) we obtain (4.40) 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 $\langle V \mid W\rangle \leqslant 0$, the inequality directly follows from Proposition 4.2, ensuring that normal vectors cannot be near-opposite. We now assume $\langle V \mid W\rangle \geqslant 0$ 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

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

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,

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

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:

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

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

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

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

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 pushforward 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 (3.12)). 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 $\|\psi - \psi_0\|_\infty \leqslant T_{tr}$ 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 $\boldsymbol{\lambda_0} + [-T_{tr}, T_{tr}]^N$ with constant

$$\varepsilon_{tr} = \delta_1 = \frac{\varepsilon_{nd}\delta_0}{2C_{\exp}C_\nabla C_{\mathrm{cond}}^2}, \tag{4.45}$$

where we recall that $\delta_0 = \varepsilon/(2^d \|\rho\|_\infty {C_{\exp}}^{2d} \operatorname{diam}(X)^d)$. 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

$$\|\hat{\rho}\|_{\mathcal{C}^{0,\alpha}(\hat{X})} \leqslant \mathrm{const}(\|\rho\|_{\mathcal{C}^{0,\alpha}}, C_{\det}). \tag{4.46}$$

We can now apply Theorem 3.1. This ensures that the function $\hat{G}$ is of class $\mathcal{C}^{1,\alpha}$ on the cube $\boldsymbol{\lambda_0} + [-T_{tr}, T_{tr}]^N$, 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 equations (4.45)–(4.46) 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 (3.13)). 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 $h := c_{y_0} - c_{y_i} = f_i \circ F^{-1}$. By definition of the push-forward, we have for any bounded measurable function χ on Ω,

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

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

$$\int_t^s \int_{f_i^{-1}(r)} \frac{\chi(F(p))\hat{\rho}(p)}{\|\nabla f_i(p)\|}\mathrm{d}\mathcal{H}^{d-1}(p)\mathrm{d}r = \int_t^s \int_{h^{-1}(r)} \frac{\chi(x)\rho(x)}{\|\nabla h(x)\|_g}\mathrm{d}\mathcal{H}^{d-1}_g(x)\mathrm{d}r. \tag{4.47}$$

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

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

depend continuously on r. Using the continuity of these functions in r, equation (4.47) and the Fundamental Theorem of Calculus, we get 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, 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, it also holds for any function χ in $\mathrm{L}^1(S)$. 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:

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

4.4. **Alternative upper transversality estimates.** Finally, we state an alternate 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

$$y \mapsto -D_x c(x', y)$$

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 the following inequality holds: for any x, $x' \in \Omega$, p_0, $p_1 \in (\exp^c_{x'})^{-1}(\Lambda)$, and $t \in [0,1]$,

$$\begin{aligned}&- c(x, \exp^c_{x'}((1-t)p_0 + tp_1)) + c(x', \exp^c_{x'}((1-t)p_0 + tp_1)))\\ &\leqslant \max\{-c(x, \exp^c_{x'}(p_0)) + c(x', \exp^c_{x'}(p_0)), -c(x, \exp^c_{x'}(p_1)) + c(x', \exp^c_{x'}(p_1))\}\end{aligned} \tag{4.48}$$

For more on these conditions, see Remark 4.3 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}.

We also slightly re-define 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* $\|\boldsymbol{\lambda}\| < T_{tr} < \frac{\varepsilon}{8C_{\exp}^{2d-2}\|\rho\|_\infty \mathcal{H}^{d-1}_g(\partial X)}$, *and* $p_0 \in K(\boldsymbol{\lambda})$ *with* $f_i(p_0) = \lambda_i$ *and* $f_j(p_0) = \lambda_j$ *for* $i \neq j$ *in* $\{1, \dots, N\}$. *Then we have*

$$\left(\frac{\langle \nabla f_i(p_0) \mid \nabla f_j(p_0)\rangle}{\|\nabla f_i(p_0)\| \, \|\nabla f_j(p_0)\|}\right)^2 \leqslant 1 - \delta_2^2 \tag{4.49}$$

where

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

The idea behind this upper bound is the following. First suppose $\nabla f_i(p_0)$ is shorter than $\nabla f_j(p_0)$ and the two vectors are collinear, and that $\boldsymbol{\lambda} = \mathbf{0}$. Inspired by calculations from [12, Remark 2.5, Proof of Lemma 4.7], this would allow us to use (4.48) to obtain the inequality $f_i \leqslant \max\{0, f_j\}$ everywhere, which in turn would mean that $-c(\cdot, y_i) + \psi(y_i)$ is less than one of $-c(\cdot, y_0) + \psi(y_0)$ or $-c(\cdot, y_j) + \psi(y_j)$ everywhere. However, combined with (Twist), this would imply that the Laguerre cell $\mathrm{Lag}_{y_i}(\psi)$ has measure zero, which is a contradiction. In the general case, we rotate $\nabla f_i(p_0)$ so that it is collinear with $\nabla f_j(p_0)$, and use the lower bound on $\mathrm{Lag}_{y_i}(\psi)$ to obtain a quantitative version of the above argument. We relegate the proof of the proposition to the Appendix in Section C.

Remark 4.3. Under a set of standard conditions, we can obtain both (QC) and (4.48).

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

- c satisfies the (Twist) condition: for every $x \in \Omega$, the map $y \in \Lambda \mapsto -D_x c(x, y)$ is a diffeomorphism onto its image $\Lambda_x := -D_x c(x, \Lambda)$ and we define the c-exponential map $\exp_x^c : \Lambda_x \to \Lambda$ by $\exp_x^c = (-D_x c(x, \cdot))^{-1}$.
- the cost $c^*(x, y) := c(y, x)$ satisfies the (Twist) condition: then for every $y \in \Lambda$, we can define the c^*-exponential map $\exp_y^c : \Omega_y \to \Omega$ by $\exp_y^c = (-D_y c(\cdot, y))^{-1}$.
- $(\exp_x^c)^{-1}(\Lambda)$ is convex for each $x \in \Omega$.
- $(\exp_y^c)^{-1}(\Omega)$ is convex for each $y \in \Lambda$.
- $\det D^2_{xy} c(x, y) \neq 0$ 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 $\eta(V) = 0$,

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

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 [22] related to regularity of optimal transport. [19, Theorem 3.2] in the Euclidean case and [16, Theorem 4.10] in the general manifold case show the above conditions imply (QC) and (4.48). In fact, they are equivalent as seen in [19]. This geometric interpretation is a key ingredient in showing regularity in the optimal transport problem in the vein of Caffarelli's classical work [5], see [10, 12].

5. Strong concavity of Kantorovich's functional

We establish in this section 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 that 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 (4.32), which can be thought of as the space of strictly c-concave functions.

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

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

where C is defined in (5.58), *and depends on $\|\rho\|_\infty$, $\mathcal{H}_g^{d-1}(\partial X)$, and $C_{\exp}$, C_∇, and ε_{tw} from Remark 4.1.*

Remark 5.1. Note that unlike the domain X, the support of the density ρ does not need be c-convex. We provide in Appendix A 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 (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 nonzero 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 (5.50). In the following, a *Lipschitz domain* denotes 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 nonempty interior since it supports an absolutely continuous probability measure. Moreover, the map $\exp_y^c$ is a diffeomorphism, hence bi-Lipschitz. This implies (i), while (ii) follows from the exact same arguments, remembering that $\rho(\mathrm{Lag}_y(\psi)) > 0$. □

Given a Lipschitz domain A of X we denote

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

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

$$\mathrm{h}(\rho) := \inf_{A \subseteq X} \frac{|\partial A|_\rho}{\min(|A|_\rho, |X \setminus A|_\rho)} \geqslant \frac{2}{C_{\mathrm{pw}}}, \tag{5.50}$$

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 [2]. Although the discussion in the reference 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. Let A be a Lipschitz domain A of $\mathrm{int}(X)$. 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 [2, 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),

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

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 used the fact that ρ is a probability measure, i.e. $\rho(X) = 1$. Proposition 10.1.2 of [2] implies that the total variation measure $|Df_n|$ narrowly converges to $|D\chi_A|$, which with the continuity of ρ implies that $\int_\Omega |Df_n|\rho \mathrm{d}\mathcal{H}^d_g$ converges to $\int_\Omega |D\chi_A|\rho \mathrm{d}\mathcal{H}^d_g = |A|_\rho$. The relation $|Df_n| = \|\nabla f_n\|_g \,\mathrm{d}\mathcal{H}^d_g$ then gives

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

Combining the previous equations together 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 [11].

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

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

The (weighted) Cheeger inequality bounds from below the first nonzero eigenvalue of the Laplacian of a weighted graph, denoted $\lambda(w)$, from its Cheeger constant and its minimal degree. The formulation we use can be deduced from Corollary 2.2 of [11] and from the inequality $1-\sqrt{1-x^2} \geqslant x^2/2$.

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

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

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

for $y \neq z$ in Y, and with zero diagonal entries ($w_{yy} = 0$). In the formula above, we used 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 nonzero 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 (4.33). By definition of the constants ε_{tw} and C_∇, for every $y \neq z$ in Y, one has

$$\varepsilon_{tw}\ w_{yz} \leqslant |\operatorname{Lag}_{y,z}(\psi)|_\rho \leqslant 2C_\nabla w_{yz}. \tag{5.51}$$

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

$$\partial A \cap \operatorname{int}(X) \subseteq \bigcup_{y\in S, z\notin S} \operatorname{Lag}_{y,z}(\psi). \tag{5.52}$$

The two inequalities (5.51) and (5.52) imply a lower bound on the numerator of the Cheeger constant:

$$|\partial A|_\rho \leqslant \sum_{y\in S, z\notin S} \left|\operatorname{Lag}_{y,z}(\psi)\right|_\rho \leqslant 2C_\nabla\ |\partial S|_w\,. \tag{5.53}$$

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

$$\mathrm{d}_y = \sum_{z\neq y} w_{yz} \leqslant \frac{1}{\varepsilon_{tw}} \sum_{z\neq y} |\operatorname{Lag}_{y,z}(\psi)|_\rho \leqslant \frac{1}{\varepsilon_{tw}} |\partial \operatorname{Lag}_y(\psi)|_\rho, \tag{5.54}$$

where the second inequality comes from the fact that the facets $\operatorname{Lag}_{y,z}(\psi)$ form a partition of the boundary $\partial\operatorname{Lag}_y(\psi) \cap \operatorname{int}(X)$ up to a $\mathcal{H}_g^{d-1}$–negligible set. To see that this is the case, 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 Lemma 3.4.

In order to apply the (continuous) Cheeger inequality, we need to replace the weighted area of the boundaries of Laguerre cells in (5.54) by the weighted volume of the cells. We have

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

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 [27, p.211]). Using the assumption $\left|\operatorname{Lag}_y(\psi)\right|_\rho \geqslant \varepsilon$, this gives us a (rather crude) reverse isoperimetric inequality

$$\begin{aligned}
\left|\partial\operatorname{Lag}_y(\psi)\right|_\rho &\leqslant \|\rho\|_\infty\, \mathcal{H}_g^{d-1}(\partial\operatorname{Lag}_y(\psi))\\
&\leqslant \frac{1}{\varepsilon} C_{\exp}^{2(d-1)}\mathcal{H}_g^{d-1}(\partial X)\left|\operatorname{Lag}_y(\psi)\right|_\rho
\end{aligned} \tag{5.55}$$

Combining (5.54), (5.55) and $|A|_\rho = \sum_{y\in S} \left|\mathrm{Lag}_y(\psi)\right|_\rho$ we obtain

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

The same inequality holds for the complement $|X \setminus S|$. We combine the previous inequality with Equation (5.53) and with Lemma 5.3 to get a lower bound on the Cheeger constant

$$\mathrm{h}(w) \geqslant \frac{\varepsilon_{tw}\varepsilon}{C_{\exp}^{2(d-1)} C_\nabla \mathcal{H}_g^{d-1}(\partial X) \|\rho\|_\infty C_{\mathrm{pw}}}. \tag{5.56}$$

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 Equation (5.55)).

Step 2. In order to apply the Cheeger inequality, we still need to bound from below the weighted degree d_y. By (5.51) one has, using the crucial fact that $\left|\partial\, \mathrm{Lag}_y(\psi)\right|_\rho$ is the measure of $\partial\, \mathrm{Lag}_y(\psi) \cap \mathrm{int}(X)$,

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

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

$$\left|\partial\, \mathrm{Lag}_y(\psi)\right|_\rho \geqslant \mathrm{h}(\rho)\ \min(\left|\mathrm{Lag}_y(\psi)\right|_\rho, \left|X \setminus \mathrm{Lag}_y(\psi)\right|_\rho) \geqslant \mathrm{h}(\rho)\varepsilon.$$

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

$$d_y \geqslant \frac{\varepsilon}{C_\nabla C_{\mathrm{pw}}}. \tag{5.57}$$

Step 3. Combining the Cheeger inequality with Equation (5.56) and (5.57) we have $\lambda(w) \geqslant C\varepsilon^3$ where

$$C := \frac{\varepsilon_{tw}^2}{2C_{\exp}^{4(d-1)} C_\nabla^3 \left(\mathcal{H}_g^{d-1}(\partial X)\right)^2 \|\rho\|_\infty^2 C_{\mathrm{pw}}^3}. \tag{5.58}$$

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 nonzero eigenvalue of the Laplacian matrix we get:

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

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 Section 4 and 5.

6.1. **General damped Newton algorithm.** Let Y be a finite set and denote $Y^{\mathbb{R}}$ the space of functions on Y. We consider $\mathcal{P}(Y)$, the space of probability measures on Y, as a subset of $Y^{\mathbb{R}}$. Finally, we denote by E_Y the space of functions on Y who 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*

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

and assume that G has the following properties:

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

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

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

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

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 $\kappa := \min(\kappa_{\frac{1}{2}\varepsilon_0}, 1)$ and $L := \max(L_{\frac{1}{2}\varepsilon_0}, 1)$. Then, the iterates (ψ_k) of Algorithm 1 satisfy

$$\begin{aligned} &\|G(\psi_{k+1}) - \mu\| \leqslant (1 - \overline{\tau}_k/2)\, \|G(\psi_k) - \mu\| \\ &\text{where } \overline{\tau}_k := \min\left(\frac{\kappa^{1+\frac{1}{\alpha}}\varepsilon}{d^{\frac{1}{\alpha}} L^{\frac{1}{\alpha}}\, \|G(\psi_k) - \mu\|}, 1\right). \end{aligned} \tag{6.59}$$

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

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

In particular, the damped Newton's 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 $\psi \in \mathcal{K}^{\frac{\varepsilon}{2}}$, the pseudo-inverse $\mathrm{D}G^+(\psi)$ maps the subspace E_Y to itself. The uniform monotonicity of G therefore implies that $\|\mathrm{D}G^+(\psi)\| \leqslant 1/\kappa$, 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 $\|v\| \leqslant \|G(\psi) - \mu\| / \kappa$. 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 $\|G(\psi_{\tau_1}) - G(\psi)\| \geqslant \varepsilon/2$, and using the Lipschitz bound on G we obtain a lower bound on τ_1

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

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

$$G(\psi_\tau) = G(\psi - \tau \mathrm{D}G(\psi)^+(G(\psi) - \mu)) = (1-\tau)G(\psi) + \tau\mu + R(\tau) \quad (6.60)$$

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

$$\begin{aligned} \|R(\tau)\| &= \left\| \int_0^\tau (\mathrm{D}G(\psi_\sigma) - \mathrm{D}G(\psi)) v \mathrm{d}\sigma \right\| \\ &\leqslant \frac{L}{\alpha+1} \tau^{\alpha+1} \|v\|^{\alpha+1} \leqslant \frac{L\|G(\psi) - \mu\|^{(1+\alpha)}}{\kappa^{(1+\alpha)}} \tau^{(1+\alpha)} \end{aligned} \quad (6.61)$$

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

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

If τ is chosen such that such that $\|R(\tau)\| \leqslant \tau\varepsilon$ we will have $G_y(\psi_\tau) \geqslant \varepsilon$ for all points 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

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

Finally we establish the second inequality required by Step 2 of the Algorithm. To do that, we subtract μ from both sides in (6.60) to obtain

$$G(\psi_\tau) - \mu = (1-\tau)(G(\psi) - \mu) + R(\tau). \quad (6.62)$$

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

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

Finally, using $L \geqslant 1$, $\kappa \leqslant 1$ and $\|G(\psi) - \mu\| \leqslant d$ (since $G(\psi)$ and μ are probability measures), we can establish that $\tau_3 \geqslant \overline{\tau}_k$ where the value of $\overline{\tau}_k$ is defined in (6.59). This ensures the first estimate on the improvement of the error between two successive steps.

By this first estimate, we see that there exists k_0 such that $\overline{\tau}_k = 1$ for $k \geqslant k_0$. When this happens, one can use (6.62) to get $\|G(\psi_{k+1}) - \mu\| \leqslant \|R(\tau)\|$. We obtain the second estimate of the theorem by plugging in (6.61). □

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 K^ε is given by

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

We have assumed that the probability density ρ is $\mathcal{C}^\alpha(X)$ 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 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 6.3). 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 uniform grid $\frac{1}{n-1}\{0,\dots,n-1\}^2$. Figure 6.3 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 [18] or [9]. It should also be possible to deal with optimal transport problems arising from geometric optics, such as the far-field reflector problem [7], whose associated cost satisfies the Ma-Trudinger-Wang condition [20].

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*

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

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.*

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

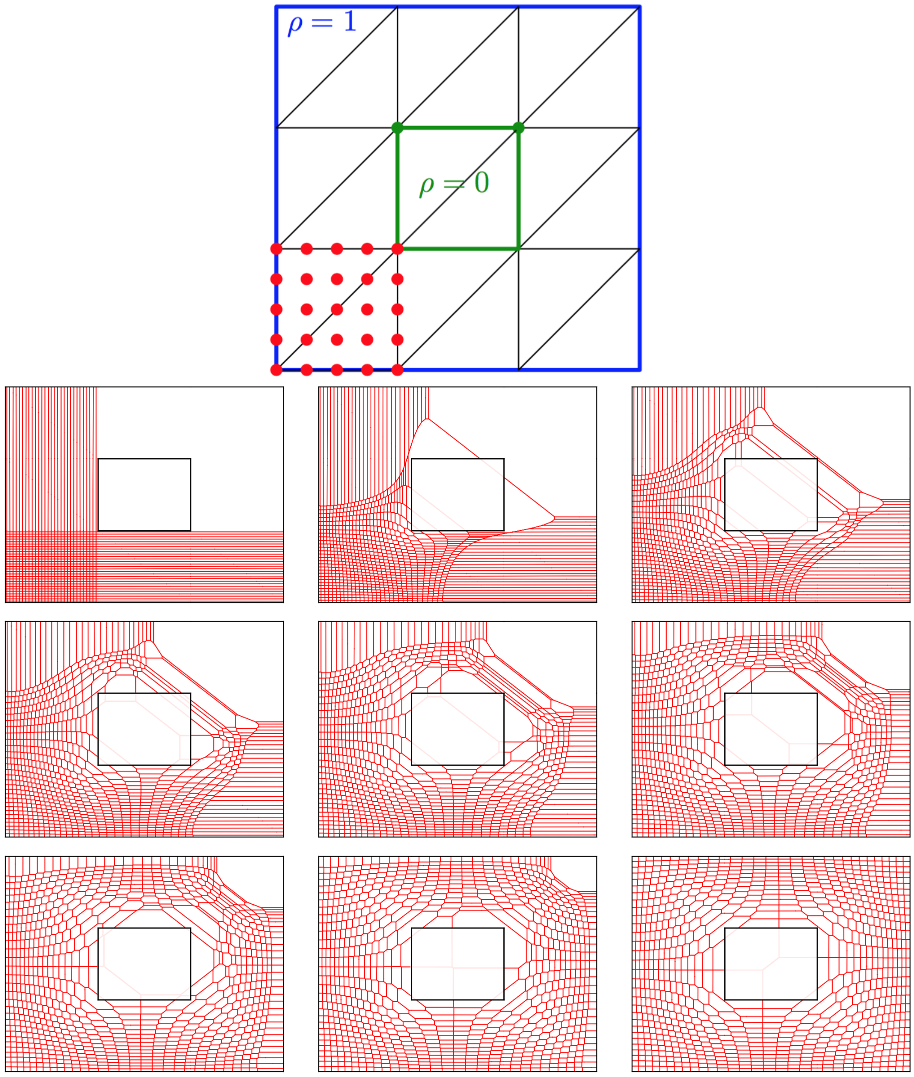

FIGURE 4. Evolution of the Laguerre cells during the execution of the damped Newton algorithm for semi-discrete optimal transport. (Top) The source density ρ is piecewise linear over the domain $X = [0, 3]^3$ over the displayed triangulation: it takes value 1 on the boundary of the square $[0, 3]^2$ 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.

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 $F_{d-1} := (1/\omega_{d-1}) \int_{\mathbb{S}^{d-1}} f(z)\mathrm{d}z$,

$$\int_{\mathbb{S}^{d-1}} |f(z) - F_{d-1}| \, \mathrm{d}\mathcal{H}^{d-1}(z) \leqslant c_d \int_{\mathbb{S}^{d-1}} \|\nabla f(z)\| \, \mathrm{d}\mathcal{H}^{d-1}(z) \qquad \text{(A.63)}$$

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

$$\int_0^R \left|\overline{f}(r) - F_1\right| \overline{\rho}(r)\mathrm{d}r \leqslant c_{\overline{\rho}} \int_0^R \left|\overline{f}'(r)\right| \overline{\rho}(r)\mathrm{d}r \tag{A.64}$$

for some positive constant $c_{\overline{\rho}}$ depending only on $\overline{\rho}$. The inequality (A.64) can be deduced from Theorem 2.1 in [1] and from the concavity of $\overline{\rho}$ on $[r,R]$.

Proof. 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 polar coordinates and the definition of ρ, one has

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

where the function $\overline{f}(r)$ 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

$$\begin{aligned} \int_{\mathrm{B}(0,R)} |f(x) - F|\,\rho(x)\mathrm{d}\mathcal{H}^d(x) &= \int_0^R \int_{\mathbb{S}^{d-1}(r)} |f(z) - F|\,\rho(z)\mathrm{d}\mathcal{H}^{d-1}(z)\mathrm{d}r \\ \leqslant \int_0^R \overline{\rho}(r)\left|\overline{f}(r) - F\right|\mathrm{d}r &+ \int_0^R \frac{\overline{\rho}(r)}{r^{d-1}\omega_{d-1}} \int_{\mathbb{S}^{d-1}(r)} \left|f(z) - \overline{f}(r)\right| \mathrm{d}\mathcal{H}^{d-1}(z)\mathrm{d}r \end{aligned} \tag{A.65}$$

We first deal with the second term of the sum. Using the Poincaré-Wirtinger inequality on the sphere (A.63), we have

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

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

$$\int_0^R \frac{\overline{\rho}(r)}{r^{d-1}\omega_{d-1}} \int_{\mathbb{S}^{d-1}(r)} \left|f(z) - \overline{f}(r)\right| \mathrm{d}\mathcal{H}^{d-1}(z) \leqslant c_d \int_{B(0,r)} \|\nabla f(x)|_{x^\perp}\| \,\rho(x)\mathrm{d}\mathcal{H}^d(x). \tag{A.66}$$

By the calculation of F above, we see 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.65) using the Poincaré-Wirtinger inequality on the segment (A.64):

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

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

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

Integrating this inequality shows that

$$\begin{aligned}\int_0^R \overline{\rho}(r) \left|\overline{f}(r) - F\right| \mathrm{d}r &\leqslant c_{\overline{\rho}} \int_0^R \frac{\overline{\rho}(r)}{\omega_{d-1} r^{d-1}} \int_{\mathbb{S}^{d-1}(r)} \left|\langle \nabla f(z) \mid \frac{z}{r}\rangle\right| \mathrm{d}\mathcal{H}^{d-1}(z) \\ &= c_{\overline{\rho}} \int_{\mathrm{B}(0,R)} \left|\langle \nabla f(x) \mid \frac{x}{\|x\|}\rangle\right| \rho(x) \mathrm{d}\mathcal{H}^d(x). \qquad \text{(A.67)}\end{aligned}$$

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

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

Using the bounds (A.66) and (A.67) in Equation (A.65), we get the desired inequality:

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

□

Appendix B. Measure of singular points on convex sets

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

$$\mathcal{N}_x K = \{v \in \mathbb{R}^d \mid \forall y \in K,\ \langle y - x \mid v\rangle \leqslant 0\}, \qquad \text{(B.68)}$$

and its elements are said to be *normal to K at x*. Let τ be a positive parameter. A point x in the boundary of K is *τ-singular* if there exist two unit vectors u, v in its normal cone $\mathcal{N}_x K$ such that $\langle u \mid v\rangle^2 \leqslant 1 - \tau^2$. Note in particular that if x is τ-singular, the linear space spanned by its normal cone has dimension two or more. Given a parameter $\tau > 0$, we consider the set of τ-singular points

$$\mathrm{Sing}(K,\tau) := \{x \in \partial K \mid \exists u, v \in \mathcal{N}_x(K) \cap \mathbb{S}^{d-1},\ \langle u \mid v\rangle^2 \leqslant 1 - \tau^2\}. \qquad \text{(B.69)}$$

The next proposition gives an upper bound on the $(d-2)$–Hausdorff measure of $\mathrm{Sing}(K,\tau)$. Note that a more general version of this result can be found in [14], with optimal constant that depends on Minkowski's quermassintegrals. We provide below a straightforward and easy proof based on the notions of packing and covering numbers.

Proposition B.1. *Let K be a convex, compact set of $\mathbb{R}^d$. Then*

$$\mathcal{H}^{d-2}(\mathrm{Sing}(K,\tau)) \leqslant \mathrm{const}(d, \mathrm{diam}(K)) \, \frac{1}{\tau},$$

where the constant depends on d and the diameter of K.

Before giving the proof of this proposition, we recall that the covering number $\mathrm{Cov}(K,\eta)$ 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

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

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

$$\mathrm{Cov}(K, \eta) \leqslant \mathrm{Pack}(K, \eta) \leqslant \mathrm{Cov}(K, \frac{\eta}{2}). \tag{B.70}$$

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

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

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 a positive radius r, we denote by K^r the set of points that are 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

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

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

$$\mathrm{Cov}(\mathcal{U}(K), \varepsilon) \leqslant \mathrm{Cov}\left(\partial K^r, \frac{\varepsilon}{L}\right).$$

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:

$$\mathrm{Cov}\left(\partial K^r, \eta\right) \leqslant \mathrm{Cov}(S, \eta) \leqslant \mathrm{const}(d) \cdot (\mathrm{diam}(K)/\eta)^{d-1}.$$

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

$$\mathrm{Cov}(U, \varepsilon) \leqslant \frac{\mathrm{const}(d, \mathrm{diam}(K))}{\varepsilon^{d-1}}. \tag{B.71}$$

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 $\langle u \mid v\rangle^2 \leqslant 1 - \tau^2$. This implies that $\mathcal{N}_x K \cap \mathbb{S}^{d-1}$ contains a spherical geodesic segment of length at least $\mathrm{const} \cdot \tau$, giving us a lower bound on the packing number of $\mathcal{N}_x K \cap \mathbb{S}^{d-1}$, namely $\mathrm{Pack}(\mathcal{N}_x K \cap \mathbb{S}^{d-1}, \eta) \geqslant \mathrm{const} \cdot \tau/\eta$. 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 $\mathrm{Card}(Y_x) \geqslant \mathrm{const} \cdot \tau/\varepsilon$. 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 $\mathrm{const} \cdot \mathrm{Card}(X) \cdot \tau/\varepsilon$. This gives

$$\mathrm{Pack}(U, 2\varepsilon) \geqslant \mathrm{const} \cdot \mathrm{Pack}(\mathrm{Sing}(K, \tau), 2\varepsilon) \cdot \tau/\varepsilon. \tag{B.72}$$

Step 3. Combining Equations (B.71), (B.72) and the comparison between packing and covering numbers (B.70), we get

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

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 $N_\varepsilon \leqslant \mathrm{const}(d, \mathrm{diam}(K))/(\tau\varepsilon^{d-2})$. By definition of the Hausdorff measure, we have

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

Appendix C. Alternate proof of upper transversality

Here we provide the proof of the alternative upper transversality estimate in Section 4.4.

Proof of Proposition 4.8. Let us again write

$$\begin{aligned} V &:= \nabla f_i(p_0),\\ W &:= \nabla f_j(p_0),\\ v &:= \frac{V}{\|V\|}, \quad w := \frac{W}{\|W\|}, \end{aligned}$$

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

$$\begin{aligned} x_0 &:= \exp^c_{y_0}(p_0),\\ q_0 &:= -D_x c(x_0, y_0), \quad q_1 := -D_x c(x_0, y_j). \end{aligned}$$

A quick calculation yields

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

Now we define the point

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

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

Thus we can define

$$\begin{aligned} y_i' &:= \exp^c_{x_0}(q'),\\ f_{i'}(p) &:= -c(\exp^c_{y_0}(p), y_i') + c(\exp^c_{y_0}(p), y_0) + c(\exp^c_{y_0}(p_0), y_i') - c(\exp^c_{y_0}(p_0), y_0) + \lambda_i, \end{aligned}$$

and by (4.48) we will obtain for all $p \in (\exp^c_{y_0})^{-1}(\Omega)$,

$$f_{i'}(p) - \lambda_i \leqslant \max\{0, f_j(p) - \lambda_j\}, \tag{C.73}$$

while another quick calculation yields

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

Now note that

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

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

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

Combining with (C.73) we then have for any $p \in (\exp^c_{y_0})^{-1}(\Omega)$,

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

or re-arranging and using that $\|\boldsymbol{\lambda}\| < T_{tr}$,

$$1 - \langle v \mid w\rangle \geqslant \sup_{p \in (\exp^c_{y_0})^{-1}(\Omega)} \frac{(f_i(p) - \max\{0, f_j(p)\} - 2T_{tr})^2}{2(C^3_{\exp} C^2_\nabla)^2}. \tag{C.74}$$

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 $X_i \cap \{f_{y_0,y_i} \leqslant -t\} \cap \{f_{y_j,y_i} \leqslant -s\}$ by

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

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

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

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

$$\begin{aligned}\mathcal{H}^d\left(X_i \cap \{-t < f_{y_0,y_i} \leqslant 0\}\right) &\leqslant \int_{-t}^0 \int_{X_i \cap \{f_{y_0,y_i} = z\}} \frac{1}{\|\nabla f_{y_0,y_i}(p)\|} d\mathcal{H}^{d-1}(p) dz \\ &\leqslant \frac{t\mathcal{H}^{d-1}(\partial X_i)}{\varepsilon_{nd}} \leqslant \frac{t C^{d-1}_{\exp} \mathcal{H}^{d-1}_g(\partial X)}{\varepsilon_{nd}},\end{aligned}$$

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 [17, Remark 5.2]. By a similar bound on the third term, we see that as long as

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

we have

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

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

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

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

Thus if we have $T_{tr} \leqslant \frac{\varepsilon\varepsilon_{nd}}{8C_{\exp}^{2d-2}\|\rho\|_\infty \mathcal{H}_g^{d-1}(\partial X)}$, combining with (C.74) we will obtain the bound (4.49) as desired. □

References

1. Gabriel Acosta and Ricardo G Durán, *An optimal poincaré inequality in* L^1 *for convex domains*, Proceedings of the American Mathematical Society (2004), 195–202.
2. Hedy Attouch, Giuseppe Buttazzo, and Gérard Michaille, *Variational analysis in sobolev and bv spaces: applications to pdes and optimization*, vol. 17, Siam, 2014.
3. F. Aurenhammer, F. Hoffmann, and B. Aronov, *Minkowski-type theorems and least-squares clustering*, Algorithmica **20** (1998), no. 1, 61–76.
4. Jean-David Benamou, Brittany D Froese, and Adam M Oberman, *Numerical solution of the optimal transportation problem using the monge–ampere equation*, Journal of Computational Physics **260** (2014), 107–126.
5. Luis A. Caffarelli, *The regularity of mappings with a convex potential*, J. Amer. Math. Soc. **5** (1992), no. 1, 99–104. MR 1124980 (92j:35018)
6. Luis A Caffarelli, Sergey A Kochengin, and Vladimir I Oliker, *Problem of reflector design with given far-field scattering data*, Monge Ampère Equation: Applications to Geometry and Optimization: NSF-CBMS Conference on the Monge Ampère Equation, Applications to Geometry and Optimization, July 9-13, 1997, Florida Atlantic University, vol. 226, American Mathematical Soc., 1999, p. 13.
7. Pedro Machado Manhaes De Castro, Quentin Mérigot, and Boris Thibert, *Intersection of paraboloids and application to minkowski-type problems*, Proceedings of the thirtieth annual symposium on Computational geometry, ACM, 2014, p. 308.
8. Fernando de Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun, *Blue noise through optimal transport*, ACM Transactions on Graphics (TOG) **31** (2012), no. 6, 171.
9. Fernando de Goes, Corentin Wallez, Jin Huang, Dmitry Pavlov, and Mathieu Desbrun, *Power particles: an incompressible fluid solver based on power diagrams*, ACM Transactions on Graphics (TOG) **34** (2015), no. 4, 50.
10. Alessio Figalli, Young-Heon Kim, and Robert J. McCann, *Hölder continuity and injectivity of optimal maps*, Arch. Ration. Mech. Anal. **209** (2013), no. 3, 747–795. MR 3067826
11. Shmuel Friedland and Reinhard Nabben, *On cheeger-type inequalities for weighted graphs*, Journal of Graph Theory **41** (2002), no. 1, 1–17.
12. Nestor Guillen and Jun Kitagawa, *On the local geometry of maps with c-convex potentials*, Calc. Var. Partial Differential Equations **52** (2015), no. 1-2, 345–387. MR 3299185
13. Alston S. Householder, *The theory of matrices in numerical analysis*, Dover Publications, Inc., New York, 1975, Reprint of 1964 edition. MR 0378371 (51 #14539)
14. Daniel Hug, *Generalized curvature measures and singularities of sets with positive reach*, Forum Mathematicum **10** (1998), no. 6, 699–728.
15. Young-Heon Kim and Jun Kitagawa, *On the degeneracy of optimal transportation*, Communications in Partial Differential Equations **39** (2014), no. 7, 1329–1363.
16. Young-Heon Kim and Robert J. McCann, *Continuity, curvature, and the general covariance of optimal transportation*, J. Eur. Math. Soc. (JEMS) **12** (2010), no. 4, 1009–1040. MR 2654086 (2011f:49071)
17. Jun Kitagawa, *An iterative scheme for solving the optimal transportation problem*, Calculus of Variations and Partial Differential Equations **51** (2014), no. 1-2, 243–263.
18. Bruno Lévy, *A numerical algorithm for* L^2 *semi-discrete optimal transport in 3d*, ESAIM M2AN **49** (2015), no. 6.

19. Grégoire Loeper, *On the regularity of solutions of optimal transportation problems*, Acta mathematica **202** (2009), no. 2, 241–283.
20. ———, *Regularity of optimal maps on the sphere: the quadratic cost and the reflector antenna*, Arch. Ration. Mech. Anal. **199** (2011), no. 1, 269–289. MR 2754343 (2011j:49081)
21. Grégoire Loeper and Francesca Rapetti, *Numerical solution of the monge–ampère equation by a newton's algorithm*, Comptes Rendus Mathematique **340** (2005), no. 4, 319–324.
22. Xi-Nan Ma, Neil S Trudinger, and Xu-Jia Wang, *Regularity of potential functions of the optimal transportation problem*, Archive for rational mechanics and analysis **177** (2005), no. 2, 151–183.
23. Quentin Mérigot, *A multiscale approach to optimal transport*, Computer Graphics Forum **30** (2011), no. 5, 1583–1592.
24. Jean-Marie Mirebeau, *Discretization of the 3d monge-ampere operator, between wide stencils and power diagrams*, arXiv preprint arXiv:1503.00947 (2015).
25. VI Oliker and LD Prussner, *On the numerical solution of the equation and its discretizations, i*, Numerische Mathematik **54** (1989), no. 3, 271–293.
26. Louis-Philippe Saumier, Martial Agueh, and Boualem Khouider, *An efficient numerical algorithm for the l2 optimal transport problem with periodic densities*, IMA Journal of Applied Mathematics **80** (2015), no. 1, 135–157.
27. Rolf Schneider, *Convex bodies: The brunn–minkowski theory*, Cambridge University Press, 1993.
28. Neil S. Trudinger and Xu-Jia Wang, *On the second boundary value problem for Monge-Ampère type equations and optimal transportation*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **8** (2009), no. 1, 143–174. MR 2512204 (2011b:49121)
29. C. Villani, *Optimal transport: old and new*, Springer Verlag, 2009.

Department of Mathematics, Michigan State University

Université Paris-Dauphine, PSL Research University, CNRS, UMR 7534, CEREMADE, 75016 Paris, France and Mokaplan, Inria

Laboratoire Jean Kuntzmann, Université Grenoble Alpes